Evaluation of cloud and water vapor simulations in CMIP5 climate models Using NASA "A-Train" satellite observations

J.H. Jiang, H. Su, C. Zhai, V.S. Perun, A. del Genio, L.S. Nazarenko, L.J. Donner, L. Horowitz, C. Seman, J. Cole, et al.

Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations

Jonathan H. Jiang,[1] Hui Su,[1] Chengxing Zhai,[1] Vincent S. Perun,[1] Anthony Del Genio,[2] Larissa S. Nazarenko,[2] Leo J. Donner,[3] Larry Horowitz,[3] Charles Seman,[3] Jason Cole,[4] Andrew Gettelman,[5] Mark A. Ringer,[6] Leon Rotstayn,[7] Stephen Jeffrey,[8] Tongwen Wu,[9] Florent Brient,[10] Jean-Louis Dufresne,[10] Hideaki Kawai,[11] Tsuyoshi Koshiro,[11] Masahiro Watanabe,[12] Tristan S. LÉcuyer,[13] Evgeny M. Volodin,[14] Trond Iversen,[15] Helge Drange,[16] Michel D. S. Mesquita,[16] William G. Read,[17] Joe W. Waters,[17] Baijun Tian,[17] Joao Teixeira,[17] and Graeme L. Stephens[17]

Received 30 November 2011; revised 2 June 2012; accepted 5 June 2012; published 18 July 2012.

[1] Using NASA's A-Train satellite measurements, we evaluate the accuracy of cloud water content (CWC) and water vapor mixing ratio (H_2O) outputs from 19 climate models submitted to the Phase 5 of Coupled Model Intercomparison Project (CMIP5), and assess improvements relative to their counterparts for the earlier CMIP3. We find more than half of the models show improvements from CMIP3 to CMIP5 in simulating column-integrated cloud amount, while changes in water vapor simulation are insignificant. For the 19 CMIP5 models, the model spreads and their differences from the observations are larger in the upper troposphere (UT) than in the lower or middle troposphere (L/MT). The modeled mean CWCs over tropical oceans range from ∼3% to ∼15× of the observations in the UT and 40% to 2× of the observations in the L/MT. For modeled H_2Os, the mean values over tropical oceans range from ∼1% to 2× of the observations in the UT and within 10% of the observations in the L/MT. The spatial distributions of clouds at 215 hPa are relatively well-correlated with observations, noticeably better than those for the L/MT clouds. Although both water vapor and clouds are better simulated in the L/MT than in the UT, there is no apparent correlation between the model biases in clouds and water vapor. Numerical scores are used to compare different model performances in regards to spatial mean, variance and distribution of CWC and H_2O over tropical oceans. Model performances at each pressure level are ranked according to the average of all the relevant scores for that level.

Citation: Jiang, J. H., et al. (2012), Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations, *J. Geophys. Res.*, *117*, D14105, doi:10.1029/2011JD017237.

1. Introduction

[2] The Intergovernmental Panel for Climate Change (IPCC) projections of climate change currently rely on some 20 climate models' simulations conducted at climate research centers worldwide. The outputs of these models consist of climate change indicators such as temperature, precipitation, clouds and water vapor. Clouds (both ice and liquid) and

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.
[2]Goddard Institute for Space Studies, New York, New York, USA.
[3]Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA.
[4]Canadian Centre for Climate Modeling and Analysis, Environment Canada, Toronto, Ontario, Canada.
[5]National Center for Atmospheric Research, Boulder, Colorado, USA.
[6]Met Office Hadley Centre, Exeter, UK.
[7]Commonwealth Scientific and Industrial Research Organisation, Clayton South, Victoria, Australia.
[8]Queensland Climate Change Centre of Excellence, Dutton Park, Queensland, Australia.
[9]Beijing Climate Center, China Meteorological Administration, Beijing, China.
[10]Laboratoire de Météorologie Dynamique, Institute Pierre Simon Laplace, Paris, France.
[11]Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan.
[12]Model for Interdisciplinary Research on Climate, Atmospheric and Ocean Research Institute, University of Tokyo, Chiba, Japan.
[13]University of Wisconsin-Madison, Madison, Wisconsin, USA.
[14]Institute for Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia.
[15]Norwegian Climate Centre, Meteorologisk Institutt, Oslo, Norway.
[16]Bjerknes Centre for Climate Research, Uni Research, Bergen, Norway.
[17]JPL, California Institute of Technology, Pasadena, California, USA.

Corresponding author: J. H. Jiang, Jet Propulsion Laboratory, California Institute of Technology, MS 183-701, 4800 Oak Grove Dr., Pasadena, CA 91109, USA. (jonathan.h.jiang@jpl.nasa.gov)

0148-0227/12/2011JD017237

water vapor, which we consider here, are important modulators of climate and are involved in feedbacks that strongly affect global circulation and energy balance. Both ice and liquid clouds significantly affect the radiation budget through their shortwave albedo and longwave greenhouse effects [e.g., *Hartmann and Short*, 1980; *Harrison et al.*, 1990; *Randall and Tjemkes*, 1991; *Bony et al.*, 2006; *Stephens*, 2005]. Water vapor produces the most important positive feedback affecting climate change [e.g., *Randall et al.* 2007; *Soden and Held*, 2006; *Hansen et al.*, 1984]. Despite all climate models producing similar magnitudes of water vapor feedback [*Randall et al.*, 2007], the simulated water vapor variabilities have large discrepancies with observations [e.g., *Pierce et al.*, 2006], and large spreads in the relation of water vapor with sea surface temperature (SST) and/or clouds [*Su et al.*, 2006a]. The uncertainties in convective parameterizations and cloud microphysics in climate models lead to uncertainties in the accuracies of simulations of water vapor and clouds and corresponding uncertainties in climate predictions. Chapter 8 of the IPCC 2007 report [*Randall et al.*, 2007] concludes that, "cloud feedbacks remain the largest source of uncertainty in climate sensitivity estimates." Improving the accuracy of cloud and water vapor simulations by climate models is thus of critical importance [e.g., *Cess et al.*, 1996; *Soden and Held*, 2006; *Bony et al.*, 2006; *Waliser et al.*, 2009].

[3] Climate modelers have, over the past decade, undertaken tremendous efforts to improve model representation of clouds and water vapor by using fine scale (large-eddy simulation or cloud-resolving) models and a variety of observations to guide their work. Many models have undergone significant changes in many areas relevant to clouds, such as the representation of the boundary layer, convection and cloud microphysics. ISCCP (International Satellite Cloud Climatology Project), ERBE (Earth Radiation Budget Experiment), SSM/I (Special Sensor Microwave/Imager), TRMM (Tropical Rainfall Measuring Mission), NVAP (NASA Water Vapor Project) and other satellite data for clouds and water vapor were used prior to 2002. The A-Train satellite constellation [*L'Ecuyer and Jiang*, 2010], which began in 2002, marks a significant improvement in observations by providing co-located and near-simultaneous 3-dimenional structures of clouds and water vapor over the globe. The A-Train observations place stringent constraints, more so than previously possible, on model simulations of clouds and water vapor, and have been used to evaluate model simulations and reanalyses data [e.g., *Li et al.*, 2005; *Pierce et al.*, 2006; *Su et al.*, 2006a; *Li et al.*, 2007, 2008; *Waliser et al.*, 2009; *Jiang et al.*, 2010; *Su et al.*, 2011; *Chen et al.*, 2011].

[4] Here, we compare multiyear means of A-Train observations with those models' results submitted to the Phase 5 of Coupled Model Intercomparison Project (CMIP5), and to their counterparts for the CMIP3. Global and zonal (tropical, midlatitude, and high latitude) multiyear spatial means and spatial distributions are considered. Special emphasis is given to vertical structure and the combined evaluation of cloud and water vapor performance. The vertical structures of clouds and water vapor are fundamentally important in determining how clouds and moisture interact with their radiative environments, precipitation and atmospheric circulation [e.g., *Kubar and Hartmann*, 2008; *Wang and Rossow*, 1998; *Holloway and Neelin*, 2009]. The model variables that we focus on are atmospheric profiles of cloud ice water content (IWC), cloud liquid water content (LWC), and water vapor mass mixing ratio (H_2O), whose evaluations over the globe were not possible prior to the A-Train era. A scoring system is devised to quantitatively evaluate and rank the CMIP5 model performances, and is applied to 30°N–30°S oceanic regions where the effects of diurnal variations are small and relevant A-Train data have best quality.

[5] When doing the comparisons we account for measurement uncertainties (including, for example, the cloud microphysical assumptions in the forward models that must be used for remote-sensing measurement retrievals), and sampling issues. Owing to extensive validation efforts on the part of instrument teams, the uncertainties (error bars) of retrievals are mostly well-defined and documented. An alternative approach for comparing model and satellite data utilizes model outputs to simulate the satellite "observables" (e.g., radiance, reflectivity, backscatter) [e.g., *Bodas-Salcedo et al.*, 2008; *Woods et al.*, 2008; *Marchand et al.*, 2009]. While such an approach has its strength, e.g., to reduce spatial and temporal sampling biases and retrieval artifacts, uncertainties of simulators are yet to be quantified (S. Klein, personal communication, 2012).

[6] The organization of the paper is as follows: section 2 describes the CMIP3/CMIP5 models and their outputs used herein; section 3 describes the A-Train data sets; section 4 compares model outputs, including differences between CMIP3 and CMIP5 model versions, and differences from the A-Train observations; and section 5 describes the scoring system and quantifies model performances based on this scoring system. An additional cloud property, cloud fraction, is discussed in the auxiliary material.[1]

2. CMIP3 and CMIP5 Climate Models

[7] The IPCC Fourth Assessment Report (AR4), released in 2007, relied heavily for climate projections on CMIP3 models. The upcoming IPCC Fifth Assessment Report (AR5) will mostly rely on the CMIP5 models. We here analyze output from 12 CMIP3 and 19 CMIP5 models that, at the time of our analyses (up to February 29, 2012), had been submitted to the Program for Climate Model Diagnosis and Inter-comparison (PCMDI) Earth System Grid (ESG) [http://pcmdi3.llnl.gov/esgcet/]. These models are listed in Table 1. Fifteen CMIP5 models are coupled atmosphere-ocean general circulation models (AOGCM), while four (CCCMA am4, GFDL am3, NCAR cam5, and UKMO hadgem2-a) are atmosphere general circulation models (AGCM).

[8] The changes in model physics from CMIP3 to CMIP5 vary from model to model. For example, in the GISS model, the rate of conversion from cloud ice to snow is increased and the influence of convectively generated snow on the glaciations of lower super-cooled liquid cloud layers is removed. From GFDL's CMIP3 cm2 to CMIP5 cm3, cloud-aerosol interaction is added and, whereas cloud particle concentrations in cm2 were specified as constants, in cm3

[1] Auxiliary materials are available in the HTML. doi:10.1029/2011JD017237.

Table 1. CMIP5 and CMIP3 Models Used in This Study

Modeling Center	Abbreviation	Model Name		CMIP5 Model Type	Resolution
		CMIP3 20c3m	CMIP5 Historical[a]		CMIP5
Beijing Climate Center, China	BCC	-	csm1.1	AOGCM	2.8125° × 2.8125°, L26
Bjerknes Centre for Climate Research;	BCCR[b]	bcm2			
Norwegian Climate Centre, Norway	NCC[b]		noresm	AOGCM	2.5° × 1.8947°, L26
Canadian Centre for Climate Modeling and Analysis, Canada	CCCMA	cgcm3.1	am4, canesm2	AOGCM AOGCM	2.8125° × 2.7673°, L35
Centre National de Recherches Météorologiques, France	CNRM	cm3	cm5	AOGCM	1.4° × 1.4°, L31
Commonwealth Scientific and Industrial Research Organization / Queensland Climate Change Centre of Excellence, Australia	CSIRO-QCCCE[c]	mk3	mk3.6	AOGCM	1.9° × 1.9°, L18
Geophysical Fluid Dynamics Laboratory, USA	GFDL	cm2	am3, cm3	AGCM AOGCM	2.5° × 2°, L48
Goddard Institute for Space Studies, USA	GISS	e-h, e-r	e2-h, e2-r	AOGCM AOGCM	5° × 5°, L29
Institute for Numerical Mathematics, Russia	INM	cm3	cm4	AOGCM	2° × 1.5°, L21
Institut Pierre Simon Laplace, France	IPSL	cm4	cm5a	AOGCM	3.75° × 1.8947°, L39
Model for Interdisciplinary Research On Climate/ Atmos. Ocean Res. Ins., U. Tokyo / Nat. Ins. Env. Std. / Japan Agency for Marine-Earth Sci. & Tech., Japan	MIROC	miroc3.2-medres[d]	miroc4h, miroc5	AOGCM	0.5625° × 0.55691°, L56; 1.4° × 1.4°, L40
Meteorological Research Institute, Japan	MRI	-	cgcm3	AOGCM	1.125° × 1.1121°, L48
National Center for Atmospheric Research, USA	NCAR	ccsm3	cesm1-cam5[e]	AOGCM	1.25° × 0.9424°, L30
Met Office Hadley Centre, UK	UKMO	hadgem1	hadgem2-es, hadgem2-a hadgem2-cc	AOGCM AGCM AOGCM	1.875° × 1.25°, L38

Modeling Center	Aerosol-Cloud Microphysics		Key References
	CMIP3	CMIP5	
Beijing Climate Center, China	No indirect aerosol effect	No indirect aerosol effect	*Wu et al.* [2010, 2012]
Bjerknes Centre for Climate Research;	No indirect aerosol effect		
Norwegian Climate Centre, Norway		Aerosol-cloud interaction	*Kirkevåg et al.* [2008] *Zhang et al.* [2012]
Canadian Centre for Climate Modeling and Analysis, Canada	No indirect aerosol effect	Aerosol 1st indirect effect	*Arora et al.* [2011]
Centre National de Recherches Météorologiques, France	No aerosol-cloud int.	Aerosol-cloud interaction	*Voldoire et al.* [2012]
Commonwealth Scientific and Industrial Research Organization / Queensland Climate Change Centre of Excellence, Australia	No aerosol-cloud int.	Aerosol-cloud interaction	*Rotstayn et al.* [2010, 2012]
Geophysical Fluid Dynamics Laboratory, USA	No aerosol-cloud int.	Aerosol-cloud interaction	*Donner et al.* [2011] *GFDL Global Atmosphere Model Development Team* [2004]
Goddard Institute for Space Studies, USA	No indirect aerosol effect	No indirect aerosol effect	*Kim et al.* [2012]
Institute for Numerical Mathematics, Russia	No indirect aerosol effect	Sulfate aerosol indirect effect	*Diansky et al.* [2002], *Diansky and Volodin* [2002] *Volodin et al.* [2010]
Institut Pierre Simon Laplace, France	Sulfate direct & 1st indirect effect	Aerosols direct & 1st indirect effect	*Dufresne et al.* [2012] *Dufresne et al.* [2005] *Hourdin et al.* [2012]
Model for Interdisciplinary Research On Climate/ Atmos. Ocean Res. Ins., U. Tokyo / Nat. Ins. Env. Std. / Japan Agency for Marine-Earth Sci. & Tech., Japan	Simple aerosol-cloud interaction	Aerosol-cloud interaction, prog. CCN	*Watanabe et al.* [2010] *Sakamato et al.* [2012]
Meteorological Research Institute, Japan	Two-moment aerosol-cloud microphysics	Two-moment aerosol-cloud microphysics	*Yukimoto et al.* [2011, 2012]
National Center for Atmospheric Research, USA	No indirect aerosol effect	Liquid & ice activation on aerosols	*Eaton* [2011] *Neale et al.* [2010]
Met Office Hadley Centre, UK	Aerosol-cloud interactions	Improved Aerosol-cloud interactions	*Martin et al.* [2011] *Collins et al.* [2011] *Jones et al.* [2011]

they are related to droplet activation that depends on aerosol properties and vertical velocity [*Ming et al.*, 2006]. Also, interactive atmospheric chemistry is added in cm3 in place of the specified chemical and aerosol concentrations in cm2 [*Donner et al.*, 2011]. The CCCMA CMIP5 differs substantially from CMIP3 in its treatment of a number of physical processes: CMIP5 includes prognostic representations of stratiform clouds; aerosol direct and indirect effects on climate; complete revision of treatments of radiative transfer, convection, and turbulent mixing. Changes to the cloud treatment in the CSIRO CMIP5 model include the coupling of warm-cloud microphysics to a prognostic aerosol scheme, a new treatment of drizzle formation (auto-conversion), and a revised treatment of the prescribed critical relative humidity for cloud formation. The latter change, in itself, could cause a substantial change in the simulated LWP compared to the CSIRO CMIP3 model. The Japanese CMIP5 miroc5 model employs an upgraded cloud parameterization scheme with more degrees of freedom than the miroc3.2 CMIP3 model. For the UKMO hadgem2, changes to the convective scheme include an "adaptive detrainment" parameterization [*Derbyshire et al.*, 2011], exponential decay of convective cloud with a half-life of 2 h, and removal of the depth criterion for shallow convection [*Gregory and Rowntree*, 1990]. Also, the treatment of aerosols in hadgem2 was improved over hadgem1, as described in *Martin et al.* [2011]. Many CMIP5 models also added at least some treatments for aerosol indirect effects that were absent in their previous CMIP3 versions (see Table 1). For example, the Russian INM cm4 model now includes the influence of prescribed sulfate aerosol concentration on cloud drop radius. For the Norwegian models, the BCCR bcm2 (CMIP3) is a different model compared to the NCC noresm (CMIP5): bcm2 is ARPEGE-based [*Déqué et al.*, 1994], whereas noresm is CCSM4-based [*Gent et al.*, 2011]. A full reference for the NCC noresm model studied here is soon to be submitted for publication. For description of processes that are central for cloud properties in the noresm, refer to *Seland et al.* [2008], *Kirkevåg et al.* [2008], and *Hoose et al.* [2009].

[9] For comparisons and evaluations, we re-grid all model data to a standard grid of 144 × 91 (longitude × latitude) with 2.5° (longitude) × 2° (latitude) horizontal resolution and 40 pressure levels from the surface to 24 hPa, with intervals of 50 hPa in the middle troposphere and finer near the boundary layer and the tropopause. The vertical interpolation is based on log-pressure. We carried out sensitivity studies and find that the different vertical interpolation methods can cause changes in computed spatial means of up to 20%, especially near the tropopause.

[10] The model results used for comparison with A-Train data are multiyear averages of the re-gridded data from the "historical*" runs for CMIP5, and the "20c3m" runs for CMIP3, which are defined as simulations of recent past climate [*Taylor et al.*, 2012]. The multiyear model averages are 20-year (1980–1999) mean when accessing changes from CMIP3 to CMIP5 (section 4); or 25-year (1980–2004) mean when comparing CMIP5 with A-Train (section 5). The different averaging periods are due to different end years of the "historical" forcings specified for CMIP3 and CMIP5.

[11] The cloud parameters in the model outputs used for this study are *clivi, clwvi, cli*, and *clw* (see the PCMDI standard output document by Karl Taylor, under "*Requested Variables*" at http://cmip-pcmdi.llnl.gov/cmip5/output_req.html). These cloud mass mixing ratio variables are monthly mean grid-box averages, taking into account clear-sky scenes and including contributions from both convective and stratiform clouds. The parameter *clivi* is the vertically integrated ice water path (IWP), *clwvi* is the vertically integrated cloud water path (CWP) that includes both IWP and liquid water path (LWP). Available in CMIP5 (not in CMIP3) are *clw*, the cloud liquid water mixing ratio, and *cli*, the cloud ice water mixing ratio, both vertically resolved. This naming convention sometimes causes confusion [e.g., *Li et al.*, 2011] since LWP should be obtained by subtracting *clivi* from *clwvi*, but LWC and IWC are obtained directly from *clw* and *cli.* The cloud water content (CWC) is the sum of *clw* and *cli.* At the time of our analysis, the *clwvi* output from the CMIP3 models BCCR bcm2 and CSIRO mk3, and the CMIP5 models CSIRO mk3.6 and IPSL cm5a, are for LWP only. We note that some modeling centers (e.g., CSIRO mk3.6) have begun to submit revisions to their data. Users are advised to check carefully the attributes in each model archive.

[12] Most models do not include snow or rain in their cloud output. The exceptions are UKMO and GFDL models. The UKMO models include snow, but not rain, in their *cli* and *clivi* and *clwvi.* For GFDL models, the inclusion or exclusion of precipitating particles depends on cloud types. Deep cumulus-generated *clw*/*clwvi* and *cli*/*clivi* include both precipitating and non-precipitating particles; shallow cumulus-generated *clw*/*clwvi* and *cli*/*clivi* do not include rain or snow; mesoscale anvils do not include liquid clouds or rain, but all sizes of ice are included in *cli*, *clivi*, and *clwvi*; for large-scale stratiform clouds, rain is not present in *clw* and *clwvi*, but all forms of precipitating and non-precipitating ice particles are included in *cli*, *clivi* and *clwvi.* We note that the inclusion of precipitating condensates in the cloud parameters by some models (e.g., GFDL and UKMO), but not all models, adds some uncertainty in our comparison work.

[13] The model parameter *prw* is vertically integrated water vapor (i.e., precipitable water), and *hus* is specific humidity. Table 2 summarizes the model output parameters used in this study.

3. A-Train Data

[14] NASA's A-Train (Aqua, Aura, CloudSat and CALIPSO satellites) carries a suite of sensors that provide nearly simultaneous and co-located measurements of

Notes to Table 1:

[a]For AOGCM, historical runs are used; For AGCM, AMIP runs (with historical forcing) are used.
[b]For the Norwegian models, the "bcm2" is developed at BCCR for CMIP3, and "noresm" was developed by NCC for CMIP5.
[c]For simplicity, acronym "CSIRO" will be used in the text for model description.
[d]For simplicity, acronym "miroc3.2" will be used in the text for model description.
[e]For simplicity, acronym "cam5" will be used in the text for model description.

Table 2. Model Outputs Used in This Study

CMIP5 Model Variable	Acronym (Unit)	Note
Ice Water Path (2D)	clivi (kg/m^2)	Mass of ice water in the column divided by area of column
Condensed Water Path (2D)	clwvi (kg/m^2)	Mass of condensed (liquid + ice) water in column divided by area of column
Mass fraction of cloud ice water (3D)	cli (kg/kg)	Mass fraction of cloud ice in atmospheric layer
Mass fraction of cloud liquid water (3D)	clw (kg/kg)	Mass fraction of cloud liquid water in atmospheric layer
Water Vapor Path (2D)	prw (kg/m^2)	Atmospheric water vapor content vertically integrated through the column
Specific humidity (3D)	hus (kg/kg)	Mass fraction atmospheric water vapor in atmospheric layer

multiple parameters that can be used for evaluating aspects of climate model performances [*L'Ecuyer and Jiang*, 2010]. The measurements used in this study, summarized in Table 3 with their estimated uncertainties, are (a) water vapor (H_2O) profiles from the Atmospheric Infrared Sounder (AIRS) onboard Aqua launched in 2002, (b) water vapor paths (WVP) from the Advanced Microwave Scanning Radiometer for Earth-Observing-System (AMSR-E) on Aqua, (c) ice/liquid water paths (IWP/LWP) from the Moderate-resolution Imaging Spectroradiometer (MODIS) on Aqua, (d) upper tropospheric H_2O and IWC profiles from the Microwave Limb Sounder (MLS) on Aura launched in 2004, and (e) LWC and IWC profiles from CloudSat launched in 2006.

[15] AIRS version 5, Level 3 H_2O product AIRX3STD is used [*Olsen et al.*, 2007]. It has spatial resolution of 50 km, but is reported on 1° × 1° (longitude × latitude) grid. The useful altitude range is 1000 hPa to 300 hPa over ocean and 850 hPa to 300 hPa over land. The estimated uncertainty is 25% in the tropics, 30% at midlatitudes, 50% at high latitudes and 30% globally averaged. These uncertainty estimates include both random and bias errors. For example, AIRS H_2O uncertainty includes the low bias because the retrievals are largely limited to clear-sky regions. The AIRS WVP over land is computed as the vertical integration of water vapor content from 850 hPa to 300 hPa and the AIRS WVP over ocean is the vertical integration from the 1000 hPa to 300 hPa.

[16] AMSR-E Level 3 WVP data of Version 5 are used [*Wentz*, 1997]. It was downloaded from the Remote Sensing Systems website (http://www.remss.com) and is reported on 0.25° × 0.25° (longitude × latitude) grids. The product is estimated to have a random error of ~1.2 kg m^{-2}. The global or tropical mean AMSR-E WVP is expected to be larger than those computed from AIRS, as AMSR-E measures the total water vapor content over the ocean from the surface to the top of atmosphere, whereas the AIRS WVP is computed as the vertical integral of water vapor content from 850 hPa to 300 hPa over land and 1000 hPa to 300 hPa over ocean. The AIRS science team has done a detailed comparison of the WVPs from AMSR-E and AIRS over ocean, and found that the difference is no more than 5% [*Fetzer et al.*, 2006].

[17] We use MODIS daily IWP and LWP data from the Collection 005 Level-3 MYD08-D3 product [*Hubanks et al.*, 2008], which were generated by sub-sampling high resolution (1 km), Level-2 swath product (MYD06) onto 1° × 1° (latitude × longitude) horizontal grids. We note that the MODIS original IWP and LWP values are for cloudy scenes only. For consistency with the gridded model data, we re-computed the MODIS original IWP and LWP to include both cloudy and clear sky scenes by multiplying the original IWP/LWP values by the cloud fractions for ice and liquid clouds, respectively. The MODIS data uncertainties mainly result from the uncertainties in the baseline and particle size distribution (PSD) assumptions. In the absence of other information, we assume a factor of 2 as a reasonable uncertainty estimate for MODIS IWP and LWP (S. Platnick, personal communication, 2011), which is similar to the IWP and LWP uncertainties described below for MLS and CloudSat.

[18] For upper tropospheric water vapor and cloud ice, we use version 2.2 Level 2 [*Livesey et al.*, 2007] MLS IWC and H_2O data sets, whose validations are described by *Read et al.* [2007] and *D. Wu et al.* [2008], respectively. These data have a vertical resolution of ~3–4 km, and horizontal resolutions of ~7 km across-track and ~200–300 km along-track. The useful altitude ranges are from 215 hPa to 83 hPa for IWC, and pressure <316 hPa for H_2O. The measurement uncertainties (including biases) for H_2O are 20% (215 hPa) to 10% (100 hPa) at tropics and midlatitudes, and ~50% at high latitude (>60°N/S) [*Read et al.*, 2007]. For IWC, there is a factor of 2 uncertainty [*D. Wu et al.*, 2008], which is mostly scaling uncertainty associated with the PSD assumption in the MLS forward model for cloud retrievals. Also MLS IWC retrieval can sometime be contaminated by gravity wave induced radiance perturbations [e.g., *Jiang et al.*, 2005] at high latitude (>45°N/S) winter, and thus only tropical to midlatitude MLS IWCs are used in this study. The MLS WVP is computed as the vertical integral of MLS H_2O from the 215 hPa to the top of atmosphere, which is added to the WVP calculated from AIRS.

[19] CloudSat IWP, LWP, IWC, and LWC from the 2B-CWC-RO (version r04) data set are used. The retrievals are

Table 3. A-Train Data Products Used in This Study

Data Source	Data Product	Acronym (Units)	Estimated Uncertainty
Aqua AIRS	Water Vapor Mixing Ratio	H_2O (g/kg)	25–30%
Aqua AMSR-E	Water Vapor Path	WVP (kg/m^2)	20%
Aqua MODIS	Ice Water Path	IWP (g/m^2)	Factor of 2
	Liquid Water Path	LWP (g/m^2)	Factor of 2
Aura MLS	Water Vapor Mixing Ratio	H_2O (ppmv)	≤20%
	Ice Water Content	IWC (mg/m^3)	Factor of 2
CloudSat	Ice Water Content	IWC (mg/m^3)	Factor of 2
	Liquid Water Content	LWC (mg/m^3)	Factor of 2

described by *Austin et al.* [2009]. These data have horizontal resolutions of ~2.5 km along-track and ~1.4 km cross-track. The vertical resolution is ~480 m, oversampled to 240 m. One of the major uncertainties is that the retrieved IWC and LWC include some contributions from precipitating particles. Thus CloudSat IWC and LWC are likely overestimated. We construct *noPcp* IWC/LWC at each grid box by removing cloud profiles where surface precipitation was detected in the grid-box average, based on the precipitation flags (rain, snow, drizzle and graupel) in the CloudSat 2C-PRECIP-COLUMN product [*Haynes et al.*, 2009]. The grid-box averages computed using all the IWC or LWC profiles are denoted as the *Total* IWC/LWC. The *noPcp* values, as noted by *Eliasson et al.* [2011], inevitably have a low bias as all "floating" ice and liquid cloud particles in addition to "falling" particles associated with precipitation events are removed in the averages. Nevertheless, the range between *noPcp* and *Total* provides a reasonable estimate of the lower and upper uncertainty bounds on CloudSat IWC and LWC. Validation studies by *Heymsfield et al.* [2008], *Eriksson et al.* [2008], and *Wu et al.* [2009], indicate that the CloudSat retrieval error is likely within ~50%. Similar to the MLS IWC, the CloudSat IWC and LWC also have uncertainty due to the PSD assumption. We estimate CloudSat IWC/LWC uncertainty to be about a factor of 2. Therefore, for the model comparisons, we use 0.5× the *noPcp* value as the lower end of the IWC/LWC, and 2.0× the *Total* value as the higher end. As CloudSat cannot accurately retrieve clouds in the lowest 1 km near the surface, we limit our analysis to cloud profiles above 1 km.

[20] All the A-Train data sets were put onto the same 144 (longitude) × 91 (latitude) × 40 (pressure) grids as done for the model outputs. The A-Train multiyear means used in evaluating the models are averages of these regridded data over the following time periods: 5 years (August 2006 to July 2010) for CloudSat; 8 years (October 2002 to September 2010) for AIRS and AMSR-E, 6 years (October 2002 to September 2008) for MODIS, and 7 years (September 2004 to August 2011) for MLS. Although the A-Train time periods do not overlap with those of the model outputs, no significant trends in clouds and water vapor are found in the model averaging periods. These multiyear means are regarded representative of "recent past climate," for which our analyses are intended.

[21] The A-Train satellites are sun-synchronous with equatorial crossings at ~1:30 pm and ~1:30 am, and this can cause sampling biases for parameters (e.g., IWC) that have diurnal variation. To reduce the effects of diurnal sampling bias, we use A-Train and model data only from the tropics and subtropics (30°N to 30°S) and only over oceanic regions when quantitatively scoring the model performances, as diurnal variations are much less over ocean than over land. We estimated the magnitude of diurnal bias in earlier versions of NCAR and GFDL models, as well as in GEOS5 reanalysis data by comparing regular modeled monthly mean IWCs with the monthly mean IWCs constructed by sampling 6-hourly model outputs onto A-Train tracks. We found that the differences between two monthly means over the tropical ocean were ~1.5% for NCAR, ~0.9% for GFDL, and ~0.1% for GEOS5 (compared to up to ~200% differences for land regions). We thus estimate that diurnal variation introduces a bias of less than 2% in the 30°N to 30°S oceanic means, significantly smaller than the measurement uncertainties. Diurnal variations over the midlatitude oceans are also relatively small, but AIRS data and wintertime MLS data have poorer quality outside the tropics. Hence, our quantitative comparison is focused on the tropical (30°N/S) oceans.

4. Comparisons of Model Outputs and A-Train Observations

4.1. IWP, LWP, and WVP

[22] Figure 1 shows the global, tropical (30°S–30°N), midlatitude (30°N/S–60°N/S) and high-latitude (60°N/S–80°N/S) multiyear averages of IWP, LWP and WVP from CMIP3, CMIP5 and A-Train. As a goal of this figure is to illustrate changes from CMIP3 to CMIP5 results, we include only models for which both CMIP3 and CMIP5 outputs were available. Grey horizontal bands in the IWP and LWP panels show the global mean 'best estimate' range - the range between CloudSat *Total* and *noPcp* global means. The factor of 2 uncertainty limits for the global mean IWP and LWP best estimates are shown by dotted lines. Note that MODIS IWPs for all three zonal means and the global mean are within the CloudSat gray band, supporting a 'best-estimate' interpretation for this band. However, MODIS provides only daytime IWP and its high-latitude mean does not include IWP from the dry polar winter. The MODIS global and midlatitude mean LWPs are within the gray band. While MODIS LWPs for tropical and high-latitude means are somewhat outside the gray band, they are within the CloudSat uncertainty range. The uncertainty limits of WVP global mean measurements, estimated as ±30% of the AIRS + MLS global mean WVP, are also shown by dotted lines. The AIRS + MLS WVPs are computed using the AIRS and MLS H_2O measurements both over land ($P \leq 850$ hPa) and over ocean. For consistency, the model WVPs are computed as the vertical integral of *hus* from 850 hPa to the top of atmosphere over land and from the surface to the top of atmosphere over ocean. The AMSR-E WVPs are the total water vapor content from the surface to the top of atmosphere, but over ocean only.

4.1.1. IWP Multiyear Global and Zonal Means

[23] The most notable change in CMIP3 to CMIP5 model outputs is the ~50% reduction of midlatitude and high-latitude IWP from GISS e-h/e-r to e2-h/e2-r, seen in the top panel of Figure 1. This reduction is largely due to the changes in the GISS model ice cloud microphysics mentioned in Section 2. Such modifications take effect mostly over the mid and high latitudes. The tropical mean IWP in GISS e2-h/e2-r is increased by ~15% compared to e-h/e-r. Although still ~30% higher than the higher end of the A-Train best-estimate, both GISS CMIP5 models produce IWP within the observational uncertainty, a significant improvement from the CMIP3 counterparts.

[24] Tropical IWP is notably increased from GFDL's CMIP3 cm2 to its CMIP5 cm3 model that implements interactive aerosols and atmospheric chemistry which were absent in the cm2. The CMIP5 models CCCMA canesm2, MIROC miroc5, and UKMO hadgem2 also show increases of global IWP from their CMIP3 counterparts, an improvement compared to the observations. However, what specific processes contributed to the improvements are not known. For the UKMO hadgem2, a recent study by *Martin et al.* [2010] have shown significant improvements globally for

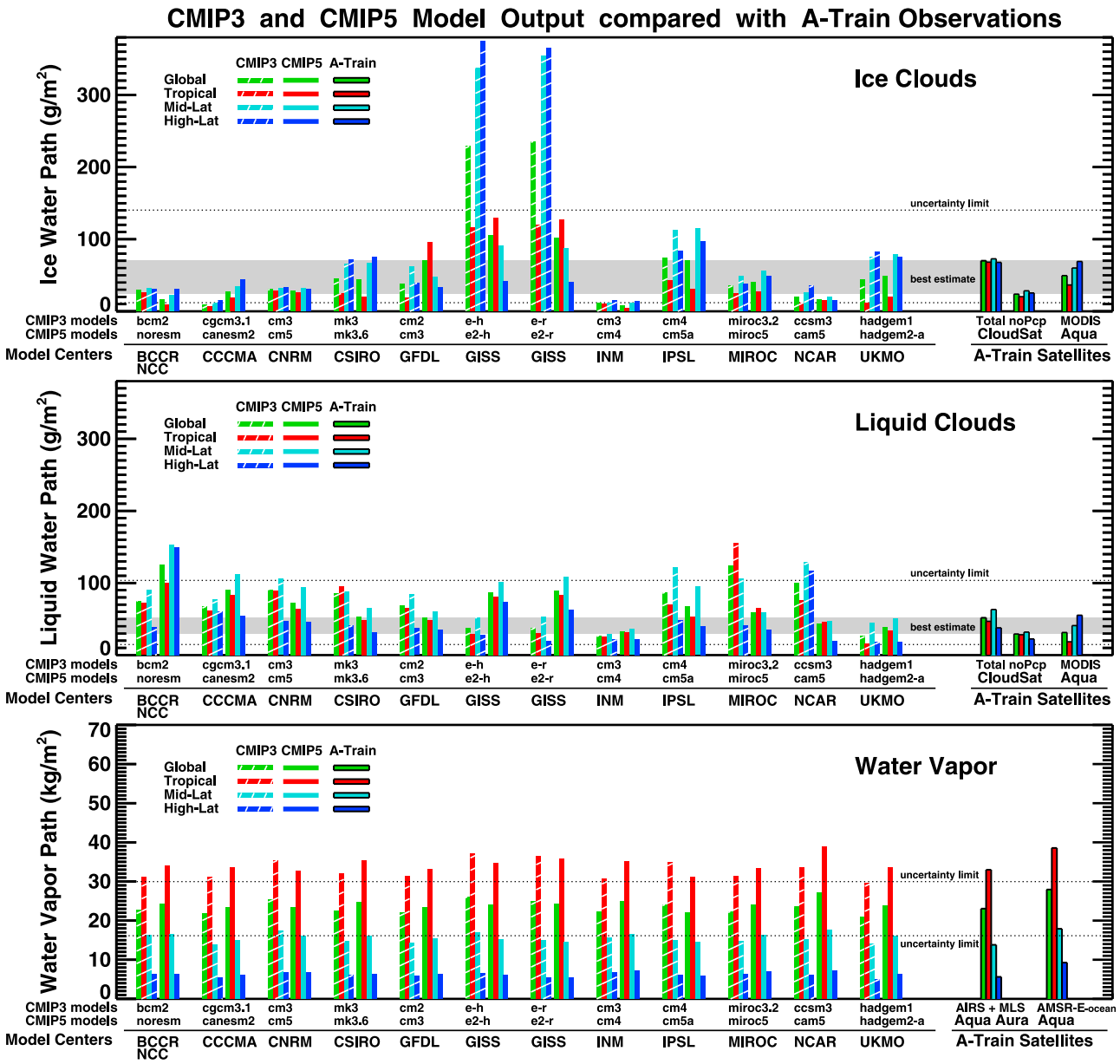

Figure 1. Multiyear mean (top) IWP, (middle) LWP, and (bottom) WVP from CMIP3 and CMIP5 models, and from A-Train observations as described in the text. Grey horizontal bands in the IWP and LWP panels show the global mean 'best estimate' range. The uncertainty limits for the global mean IWP and LWP best estimates are shown by dotted lines. In Figure 1 (bottom), WVPs from AIRS + MLS and from the models are computed from 850 hPa to the top of atmosphere over land and from the surface to the top of atmosphere over ocean. The uncertainty limits for the AIRS + MLS global mean WVP are shown by dotted lines. The AMSR-E WVPs are the total water vapor content from the surface to the top of atmosphere, but over ocean only.

the simulation of cloud amount and humidity compared to its predecessor hadgem1. This is particularly apparent in the tropics and results primarily from the changes to the convection scheme.

[25] Reductions of IWP in CMIP5 compared to CMIP3 are seen in INM cm4 and NCAR cam5. The NCC noresm in CMIP5 also has smaller IWP compares to the BCCR bcm2 in CMIP3. The IPSL cm5a model is very similar to the previous IPSL cm4 model except for the improvements in horizontal and vertical resolutions [*Dufresne et al.*, 2012] and little change is shown in its IWP. The CNRM and CSIRO models also have little changes in IWP from CMIP3 to CMIP5.

[26] Overall, of the 12 model pairs examined, 7 CMIP5 IWPs are within the CloudSat "best estimate" gray band, and 11 (all except INM cm4) are within the observational uncertainty limits. This is an improvement over CMIP3, where 6 models have IWPs within the gray band and 8 have IWPs within the uncertainty limits.

4.1.2. LWP Multiyear Global and Zonal Means

[27] Figure 1 (middle) shows LWP for all models and A-Train observations. Increases from CMIP3 to CMIP5 model outputs are seen in CCCMA canesm2, GISS e2-h and e2-r, INM cm4, and UKMO hadgem2. The NCC noresm in CMIP5 also has much larger LWP compares to the BCCR bcm2 in CMIP3. Reductions in LWPs from CMIP3 to CMIP5 are seen in CNRM cm5, CSIRO mk3.6, GFDL cm3, IPSL cm5a, NCAR cm5, and MIROC miroc5. Some of these changes in LWP are related to changes in cloud treatment in the models. For example, the CSIRO model includes a simple treatment of sub-grid moisture variability, in which the width of sub-grid moisture distribution is parameterized via a prescribed critical relative humidity (RH_c) for onset of cloud formation [*Rotstayn*, 1997]. In CSIRO mk3, RH_c would decrease between cloud base and top in convective columns when convection occurs. It was shown that such RH_c treatment leads to an increase of LWP, while IWP is relatively insensitive to RH_c [*Rotstayn*, 1999]. This RH_c reduction was removed in CSIRO's mk3.6, in which the RH_c is prescribed and not dependent on convection. Such change explains a substantial decrease of LWP from mk3 to mk3.6, in conjunction with a relatively small change in IWP.

[28] Global mean LWPs within the gray band are produced by 4 CMIP5 models: GFDL cm3, INM cm4, NCAR cam5, and UKMO hadgem2. Eleven CMIP5 models (all except NCC noresm) have LWPs within the observational uncertainty. In contrast, only 2 CMIP3 models (GISS e-h and e-r) yield global mean LWPs within the gray band, and 11 CMIP3 models (all except MIROC miroc3.2) have LWPs within the observational uncertainty.

4.1.3. WVP Multiyear Global and Zonal Means

[29] Figure 1 (bottom) shows WVP. Model differences are within ~10%, and changes from CMIP3 to CMIP5 are less than 5%. The differences between model and AIRS + MLS observations are less than ~15%, well within the 30% observational uncertainty. The difference between AIRS + MLS and AMSR-E are mainly due to the fact that AMSR-E WVPs do not include data over land, whereas the AIRS + MLS (and all models') WVPs are averaged using data over both ocean (pressure $\leq$ 1000 hPa) and land (pressure $\leq$ 850 hPa).

4.1.4. IWP Multiyear Mean Spatial Distributions

[30] Figure 2 shows the multiyear mean spatial distributions of IWP from the CMIP3 and CMIP5 models and from the A-Train. The corresponding Taylor Diagram for IWP is displayed in Figure 5 (top). Of the 12 CMIP5 models examined, comparisons with the observations indicate that 6 models (CCCMA canesm2, GFDL cm3, GISS e2-r/e2-r, MIROC miroc5, and UKMO hadgem2-a) show IWP improvements from CMIP3, 3 show little change (CNRM cm5, CSIRO mk3.6, and NCAR cam5), and 2 appear degraded (IPSL cm5a and INM cm4). CMIP5 NCC noresm also perform poorer than CMIP3 BCCR bcm2 in simulating IWP.

[31] The IWP Taylor diagram (Figure 5, top) suggests that there is a large spread among the model simulated standard deviations - from as small as 0.05× to as large as 4.5× the observed. The most significant improvements from CMIP3 to CMIP5 are found in the two GISS models (e2-h/e2-r), in which substantial reduction in mid and high latitude and increase in the tropics result in better agreement with the observations, reducing the RMS errors from 4.5 to less than 2 and improving spatial correlations from ~0.35 to ~0.5. The GFDL cm3 has IWP increase in the tropics but decrease in the northern hemispheric storm tracks and southern mid and high latitudes, yielding better agreement with observations in the tropics, but a low bias in the mid and high latitudes. Overall, the GFDL cm3 is improved over its previous cm2 as the spatial correlation to the observation increases from 0.35 to over 0.6. For CCCMA's and MIROC's CMIP5 models, the IWPs are increased slightly over both the tropics and midlatitudes, bringing the standard deviations slightly closer to the observed. For the UKMO hadgem2-a, there is a slight increase in IWP in the tropics, associated with smaller RMS errors. Its IWP has little changes in the mid- and high latitudes.

[32] CNRM cm5, CSIRO mk3.6 and NCAR cam5 all show very little change in IWP and no obvious improvements from CMIP3 to CMIP5. For other three CMIP5 models, NCC noresm has overall reduction in IWP comparing to BCCR bcm2 in CMIP3, resulting in a low bias compared to the observations. INM cm4 has IWP decreased in the equatorial eastern Pacific but increased over the midlatitude storm tracks. The global mean is not significantly changed, but there is a noticeable degradation in the agreement with observations over the inter-tropical convergence zone (ITCZ). The changes in IPSL cm5a are small, but the slight reduction in IWP in the tropics results in a slight degradation as reflected in reduced correlation on the Taylor diagram.

[33] In terms of spatial correlation and standard deviation, the multimodel mean IWP for CMIP5 shows a substantial improvement from CMIP3: the RMS is reduced from 1.03 to 0.65. The spreads between models are much larger than the respective CMIP5 and CMIP3 differences, except the two GISS models that exhibit substantial improvements from CMIP3 to CMIP5.

4.1.5. LWP Multiyear Mean Spatial Distributions

[34] Figure 3 shows the multiyear mean spatial distributions of LWP from the CMIP3 and CMIP5 models and from the A-Train, with the corresponding Taylor Diagram (Figure 5, middle). Of the 12 models examined, 7 show LWP improvements from CMIP3 to CMIP5, 3 show changes but no notable improvements, while 2 appear degraded, compared with the observations.

[35] The models with improved agreement include CNRM cm5, CSIRO mk3.6, GFDL cm3, INM cm4, IPSL cm5a, MIROC miroc5, and NCAR cam5. From the LWP Taylor diagram (Figure 5, middle), we can see the improvements of CNRM, CSIRO, IPSL, and MIROC models in all parameters: better standard deviation and correlation, and smaller RMS errors. For CNRM, the LWP values are reduced slightly from cm3 to cm5, resulting in slightly improved agreement with the observations. For CSIRO, LWPs are reduced in midlatitudes, corresponding to substantial improvement (in both amount and distribution). Also notable is the improved simulation of clouds in the eastern Pacific subsidence region and the southern Indian Ocean. Substantial LWP reduction is also seen in miroc5, leading to better agreement with the observations. For IPSL, LWPs in cm5a are slightly reduced in both the tropics and midlatitudes comparing to cm4, a better agreement with the observations. The improvements in GFDL and NCAR models are indicated by substantial reduction in RMS errors - their CMIP3 and CMIP5 models

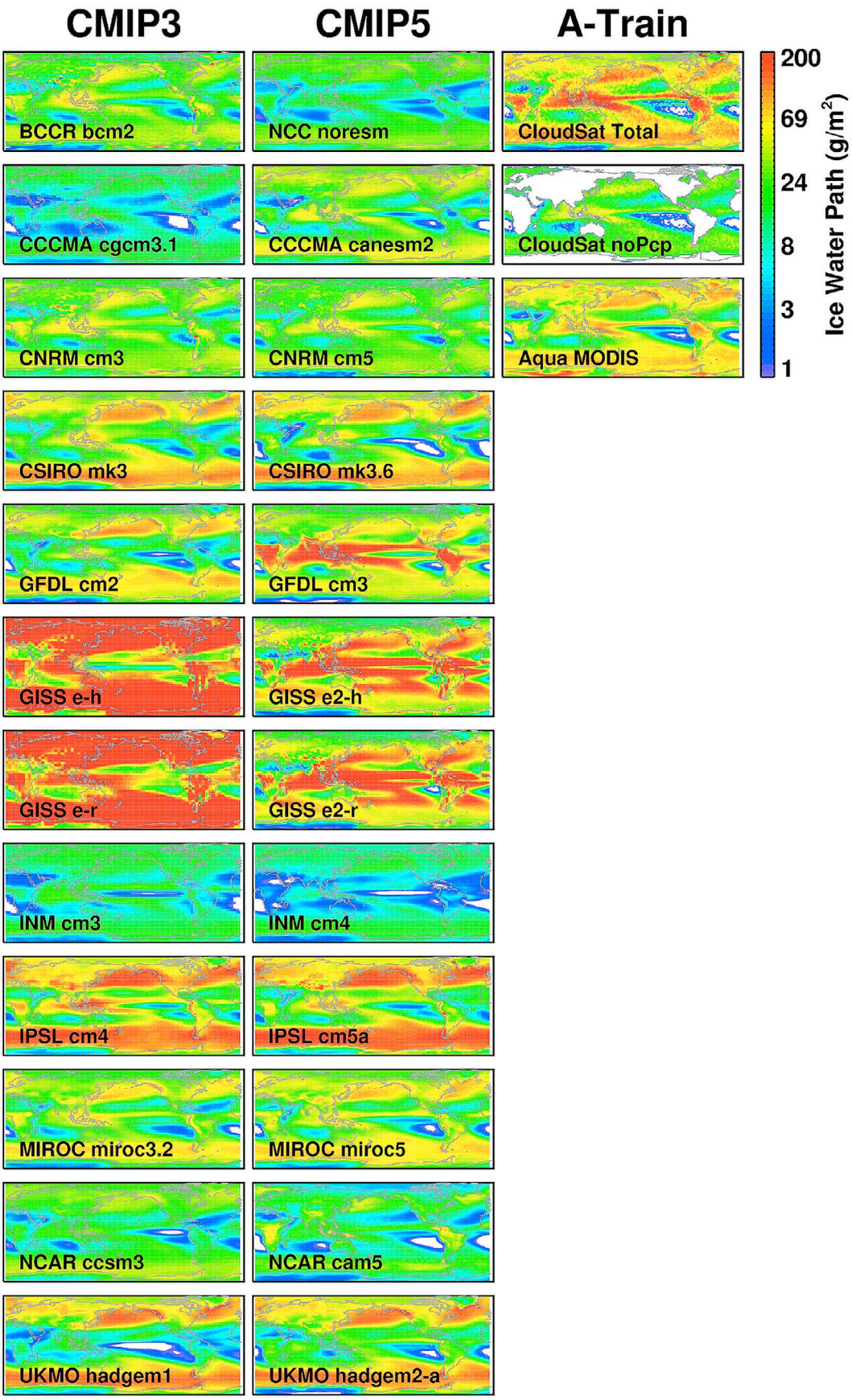

Figure 2. Multiyear mean IWP from CMIP3 and CMIP5 models, and from A-Train observations.

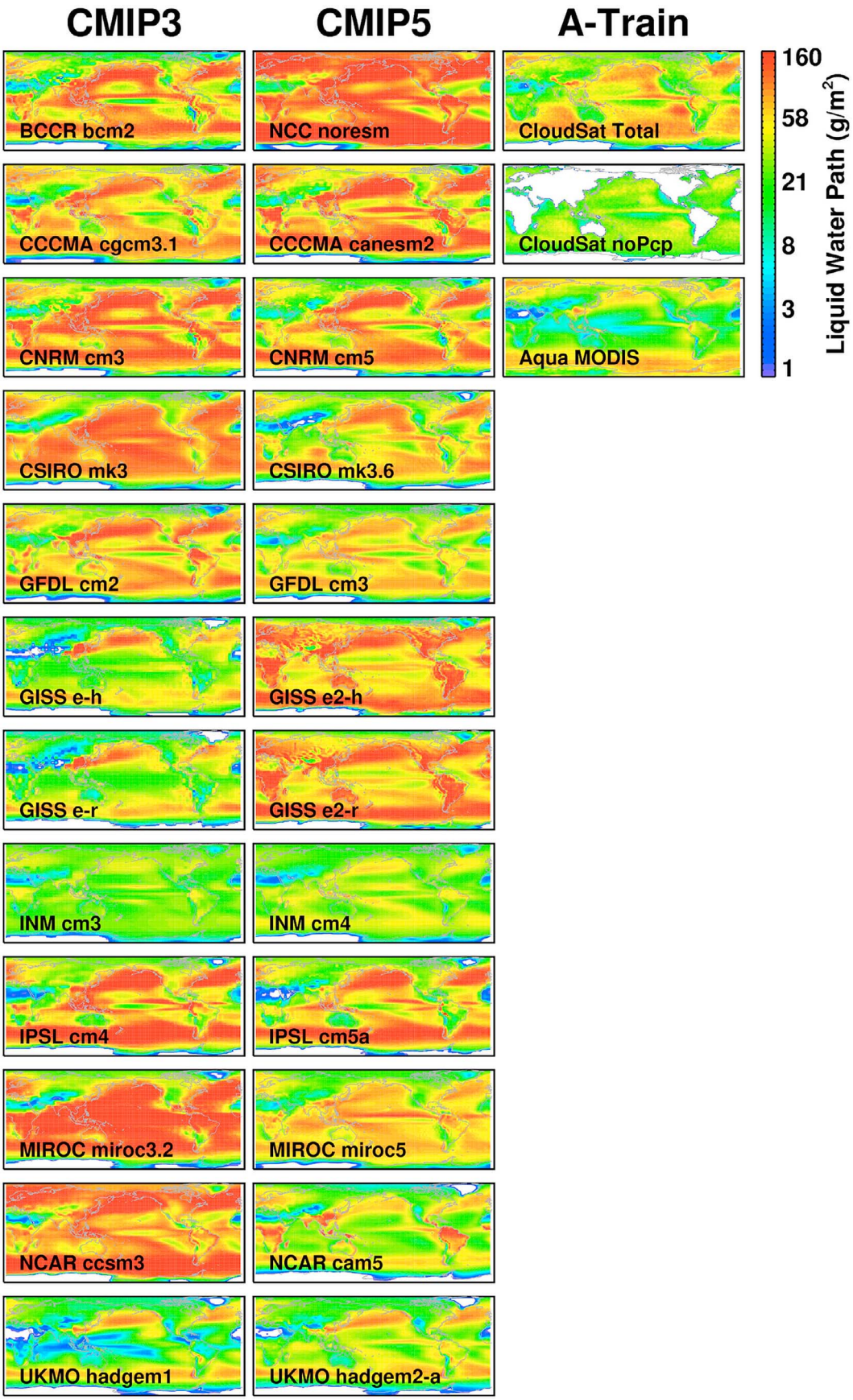

Figure 3. Multiyear mean LWP from CMIP3 and CMIP5 models, and from A-Train observations.

have similar spatial patterns, but magnitude of LWP is reduced. For INM cm4, the standard deviation is closer to the observation than the previous cm3.

[36] The degraded models are NCC noresm and CCCMA canresm, indicated by much larger RMS errors compared to the observation. Both models have large increase in LWP, which leads to significant overestimate compared to the observation, worse performance than their CMIP3 counterparts. For CCCMA canesm2, the appearance of a "double ITCZ" in the equatorial Pacific also contributes to the poorer agreement with observations.

[37] The two GISS models e2-h/e2-r and the UKMO hadgem2-a show increased LWPs, but no obvious improvement or degradation from CMIP3 to CMIP5.

[38] The LWP multimodel mean for CMIP5 has the same spatial correlation as that for CMIP3, around 0.5, while the RMS for multimodel mean is reduced somewhat from 0.89 for CMIP3 to 0.86 for CMIP5. The model differences between CMIP3 and CMIP5 counterparts are noticeably small compared to the spread among models.

4.1.6. WVP Multiyear Mean Spatial Distributions

[39] Figure 4 shows the multiyear mean spatial distributions of WVP from the CMIP3 and CMIP5 models and from the A-Train. From this figure and the Taylor diagram shown in Figure 5 (bottom), we can see there is overall good agreement with the observation, and model differences are small. Since the variability of WVP is dominated by lower-tropospheric water vapor, it is expected that the simulated lower-tropospheric water vapor is similar among models, while large discrepancy may exist in the upper troposphere as we will discuss later. The multimodel mean for CMIP5 is slightly better than that for CMIP3, with slightly reduced RMS error from 0.20 to 0.17. The spatial correlation is about 0.98.

4.2. Vertical Profiles of CWC, IWC and H_2O

[40] Figure 6 shows the multiyear mean vertical profiles of CWC and IWC (Figure 6, top) and H_2O (Figure 6, bottom) from the 19 CMIP5 models and from the A-Train observations. The 'best estimated' CWC values from the CloudSat observations are indicated by the gray band between the CloudSat *noPcp* and *Total* values. Observational uncertainty limits are indicated by the dotted lines. There is a large spread among model CWC in all three latitude bands and globally. At 300 hPa, for example, the global mean CWC from GISS e2-r is more than 200× larger than from INM cm3. The modeled tropical CWCs range from ∼3% to ∼15× of the MLS IWC in the upper troposphere. For mid-troposphere 700 hPa to 400 hPa, the modeled tropical CWCs are from ∼30% to ∼4× of the CloudSat *Total.* In lower troposphere, the modeled CWCs are ∼40% to 2× of the CloudSat *Total.*

[41] H_2O (Figure 6, bottom) differences among the models are within 20% in the mid- and lower troposphere, but more than 400% above ∼200 hPa altitude. Model differences from the AIRS observations are small (<10%) in the mid- and lower troposphere, but range from ∼1% to ∼200% of the MLS observations at 100 hPa. All models are biased high compared to AIRS observations in the mid- and lower troposphere in all latitude bands, but mostly within the observational uncertainty. Relative to MLS observations, most models are biased high between 300 and 120 hPa in the tropics and midlatitudes, and between 300 and 150 hPa in the high latitudes; The biases can be larger than the MLS uncertainty. Above ∼120 hPa altitude in the tropics and midlatitudes and ∼150 hPa altitude in the high latitudes, the model biases can be either positive or negative. Figure 7 shows the ratio of modeled H_2O to AIRS and MLS observations as a function of height, which further demonstrates that the inter-model spread in percentage is larger in the upper troposphere comparing to the mid- and lower troposphere.

[42] Figure 8 shows the multiyear zonal means of CWC and H_2O as a function of latitude and height. We notice that all models generally underestimate IWC in the tropical upper troposphere but produce reasonable amounts of CWC in the extra-tropics. This might be because high-level ice clouds are generally associated with synoptic uplift, which is resolved in the models, whereas in the tropics they often result from convective detrainment more difficult to simulate. Not counting snow in IWC in the models may also contribute to the underestimate of upper-tropospheric IWC.

[43] All models produce similar zonal mean distributions of water vapor. The major differences from observations are in the upper troposphere as discussed in the following sections.

5. Quantitative Evaluation of Model Performances

[44] In this section we quantify the differences between model and A-Train multiyear means, and score the model performances compared to the observations. We focus on 30°S–30°N oceanic regions in this study, where the A-Train data has best quality and diurnal sampling bias is relatively small.

5.1. The Scoring System

[45] Model performance is evaluated with a system that scores how well each model multiyear mean reproduces the A-Train multiyear mean in terms of (1) spatial means, (2) spatial variances, and (3) spatial distributions. Our scoring system follows that of *Douglass et al.* [1999], *Waugh and Eyring* [2008], and *Gettelman et al.* [2010], but with additional considerations of observational uncertainties.

[46] We define the spatial mean scores G_m for IWC, LWC and H_2O as

$$G_m^{IWC,LWC} = \max\left[0,\ 1 - \frac{1}{n_g}\frac{\left|\ln\left(m_{mdl}^{IWC,LWC}\right) - \ln\left(m_{obs}^{IWC,LWC}\right)\right|}{\ln \varepsilon_{m,obs}^{IWC,LWC}}\right], \tag{1}$$

$$G_m^{H_2O} = \max\left[0,\ 1 - \frac{1}{n_g}\frac{\left|m_{mdl}^{H_2O} - m_{obs}^{H_2O}\right|}{\varepsilon_{m,obs}^{H_2O}\, m_{obs}^{H_2O}}\right], \tag{2}$$

where m denotes the 30°N–30°S oceanic spatial mean, *mdl* denotes model value, *obs* denotes observational value, and $\varepsilon_{m,obs}$ is the fractional uncertainty of the observed spatial mean. The observed IWC and LWC spatial means have a factor of 2 uncertainty; hence $\varepsilon_{m,obs}^{IWC,LWC} = 2$. The H_2O observational uncertainties $\varepsilon_{m,obs}^{H_2O}$ are 0.1 at 100 hPa, 0.2 at

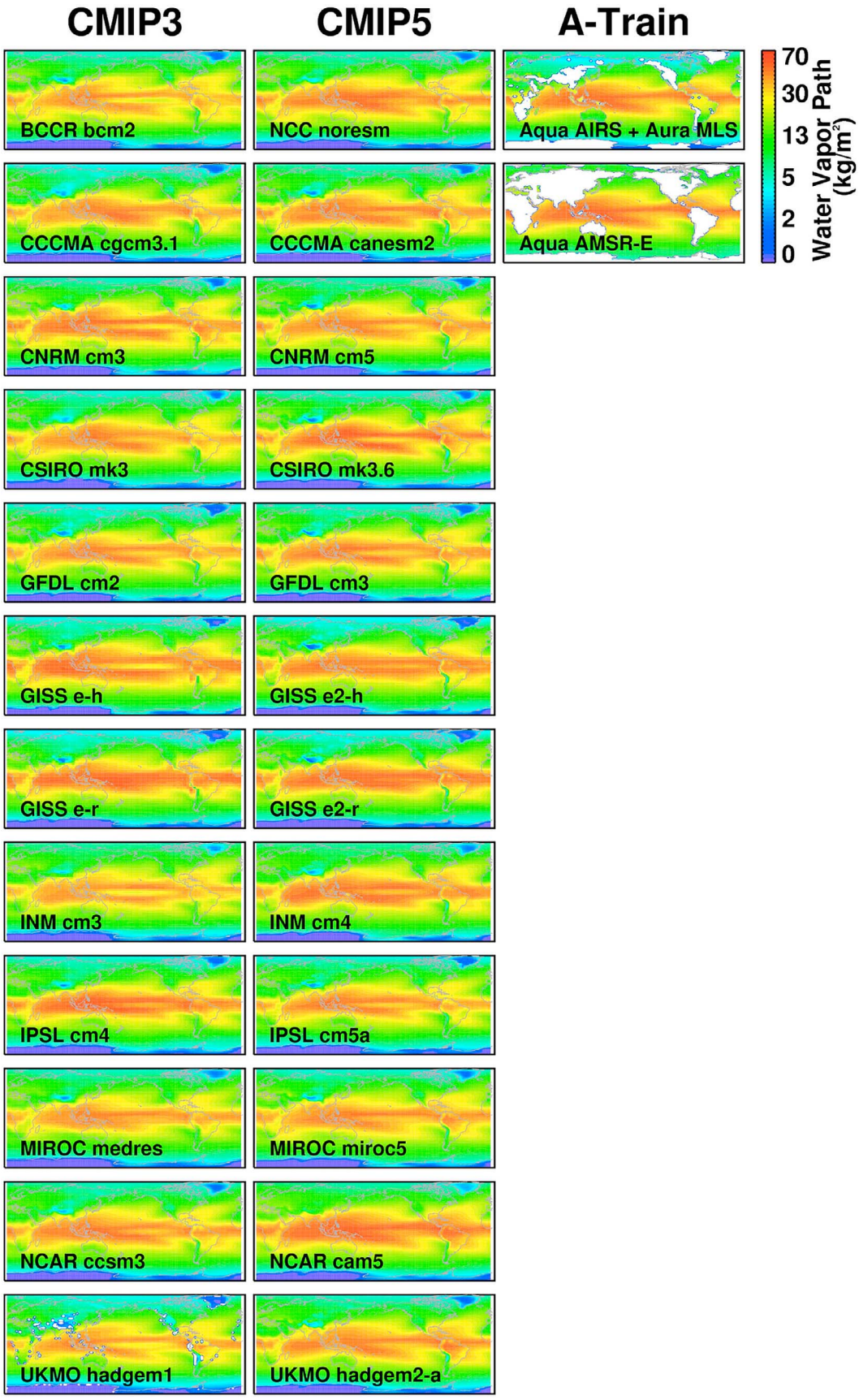

Figure 4. Multiyear mean WVP from CMIP3 and CMIP5 models, and from A-Train observations.

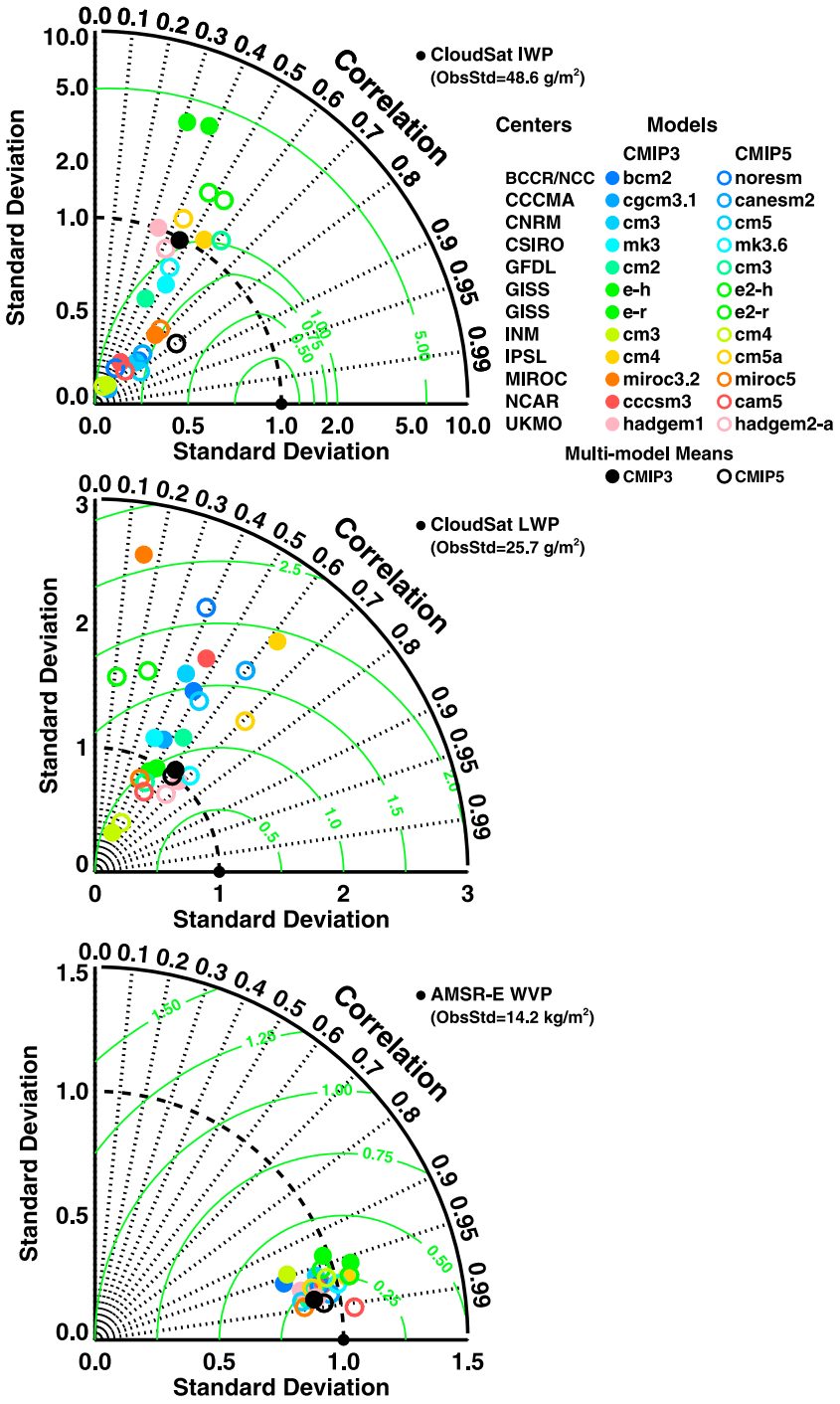

Figure 5. Taylor diagrams showing the global (80°N–80°S) oceanic multiyear mean IWP, LWP, and WVP simulations from the CMIP3 and CMIP5 models (colored symbols) as compared to the A-Train observations (the black dot on the horizontal axis with the value of 1 = the standard deviation of the observed variable). The horizontal axis represents the fraction of the modeled spatial variation pattern that can be explained by the observed spatial pattern. The vertical axis represents the standard deviation of the modeled spatial pattern orthogonal to the observation, which is normalized by the observed standard deviation. The distance to the origin from each point in the Taylor Diagram corresponds to the spatial standard deviation of modeled variable and the distance of each point to the observed point (1, 0) on the *x* axis is the RMS of the difference between the modeled and observed quantities, as scaled by the green arc-lines. The correlation between the modeled and observed quantities is marked by the numbers on the black arc. Note due to the large spread of the modeled IWPs, a mixed linear-log scale is used for the vertical and horizontal axes.

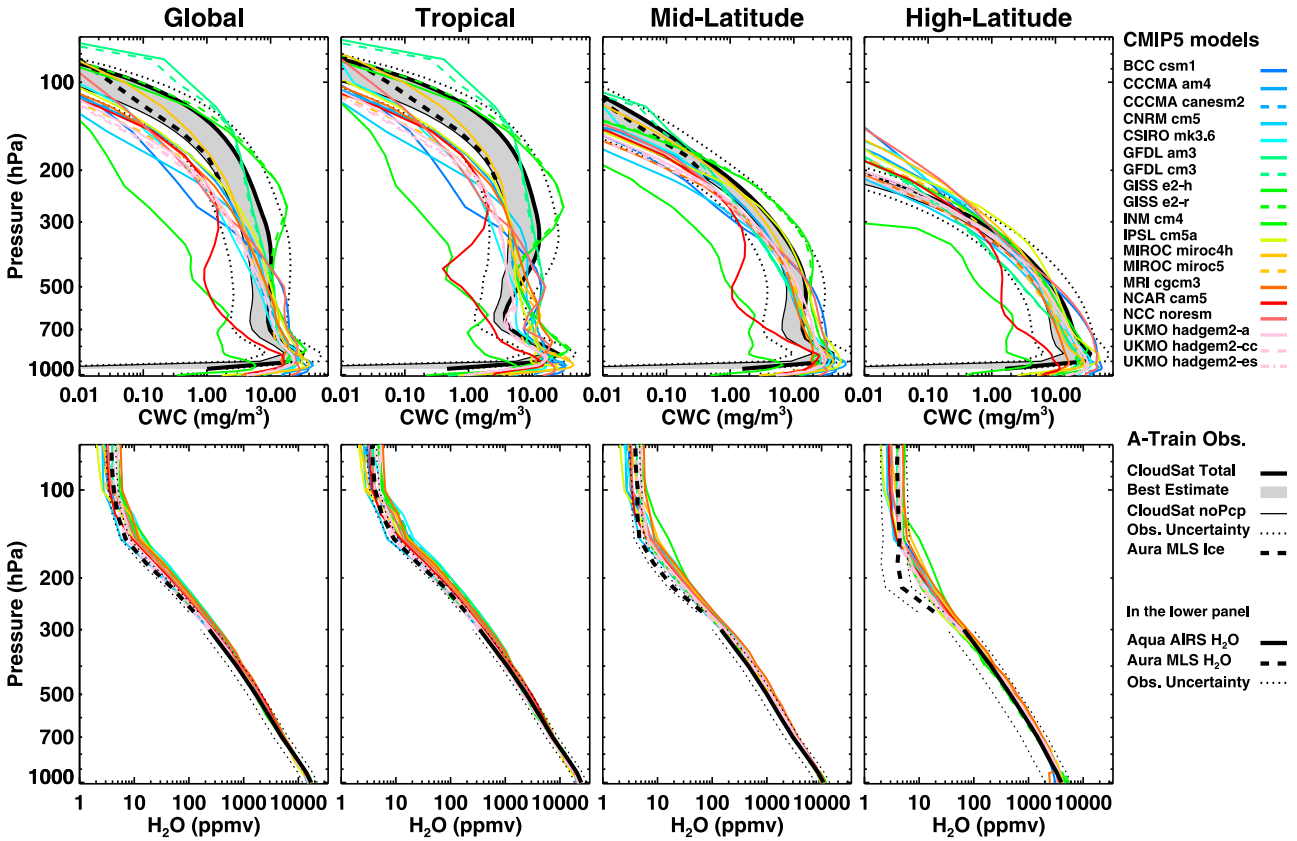

Figure 6. Multiyear mean CWC (upper-panels) and H_2O (lower-panels) vertical profiles from CMIP5 models and from A-Train observations. In the upper-panels, the 'best estimated' CWC values from the CloudSat observations are indicated by the gray band between the CloudSat *noPcp* and *Total* values. The CWC observational uncertainty limits are indicated by the dotted lines. MLS IWC profiles, plotted for $P \leq 215$ hPa, are located in the best-estimated zone, as expected. In the lower-panels, H_2O data from Aqua AIRS are available at and below 300 hPa altitude and H_2O from Aura MLS are available above the 300 hPa altitude. The H_2O observational uncertainty limits are also shown by the dotted lines.

215 hPa, and 0.25 at 600 and 900 hPa. The scaling factor n_g is chosen to be 3, except for LWC at 900 hPa where $n_g = 4$ is chosen to account for a greater uncertainty (e.g., larger difference between *Total* and *noPcp*) in LWC there. Due to the large range of values, the difference in logarithms is used for IWC and LWC. In this grading system, for example, a zero G_m score means: (1) for H_2O, the model-observation difference is greater than 3× the observational uncertainty, and (2) for IWC/LWC, the model value is either 8× greater (16× for 900 hPa) or less than 1/8 (1/16 for 900 hPa) the observational value.

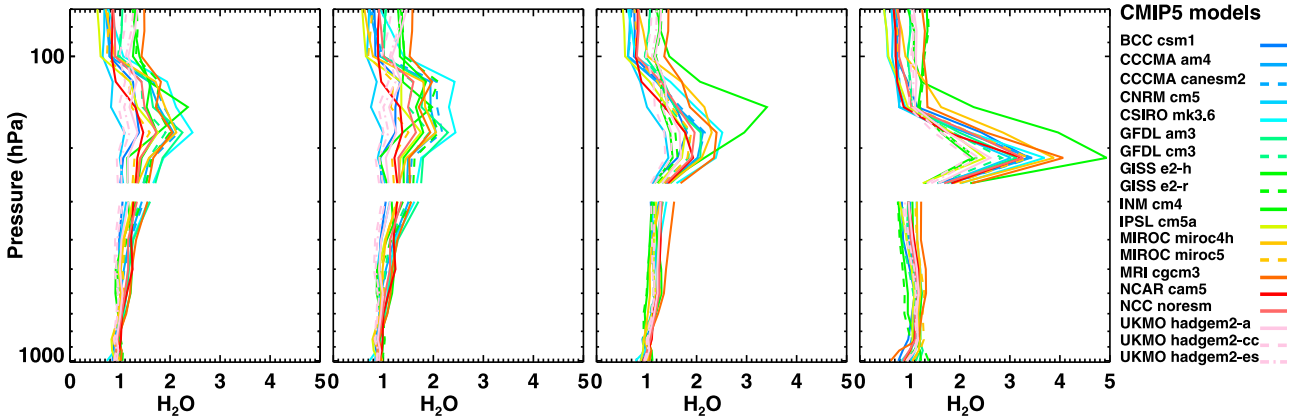

Figure 7. Ratio of multiyear CMIP5 modeled H_2O to A-Train observed values as a function of height.

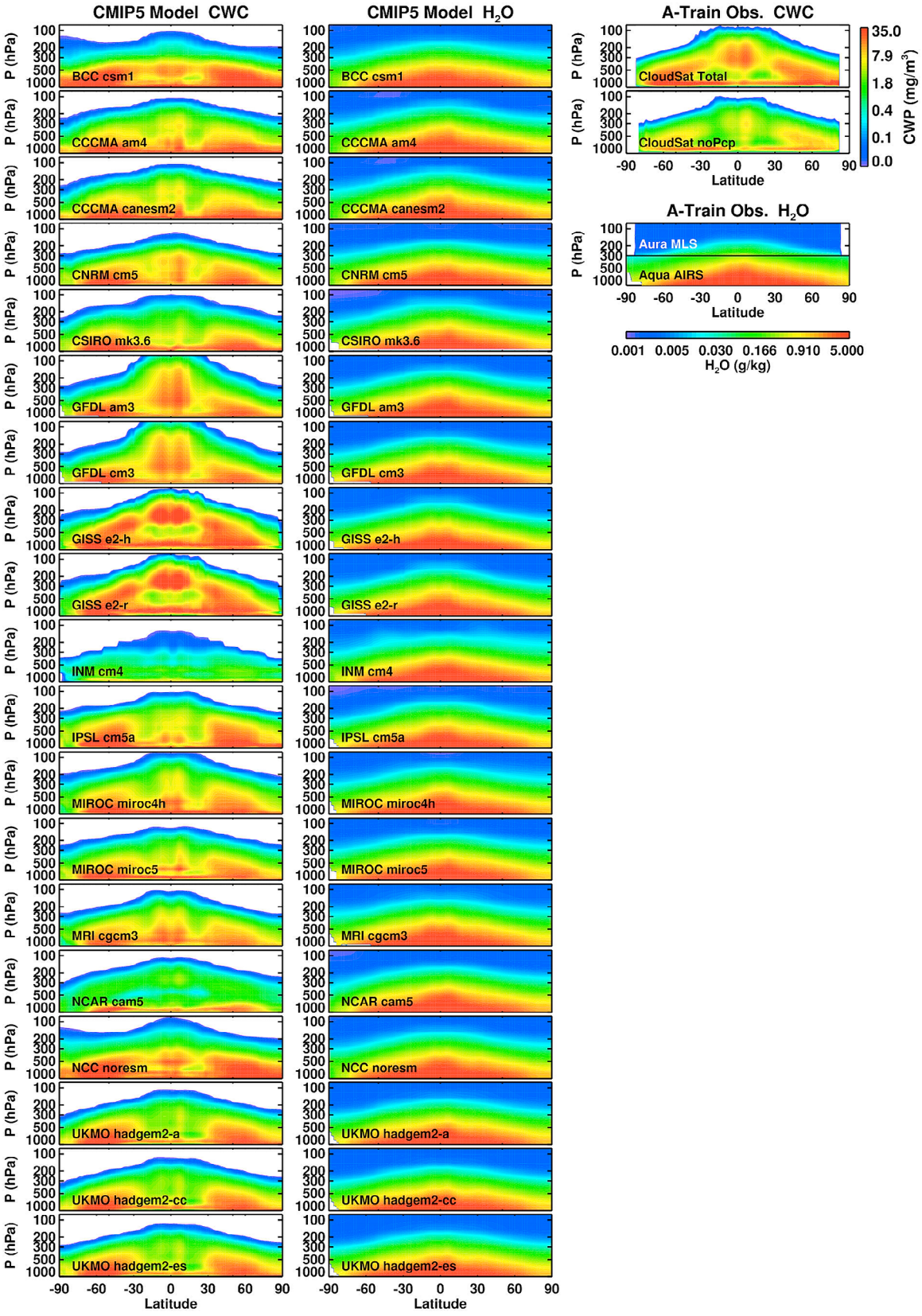

Figure 8. Multiyear mean zonal profiles of CWC and H_2O from CMIP5 models and from A-Train observations. For Aura MLS observation, H_2O is plotted for $P < 300$ hPa, and for Aqua AIRS observation, H_2O is plotted for $P \geq 300$ hPa.

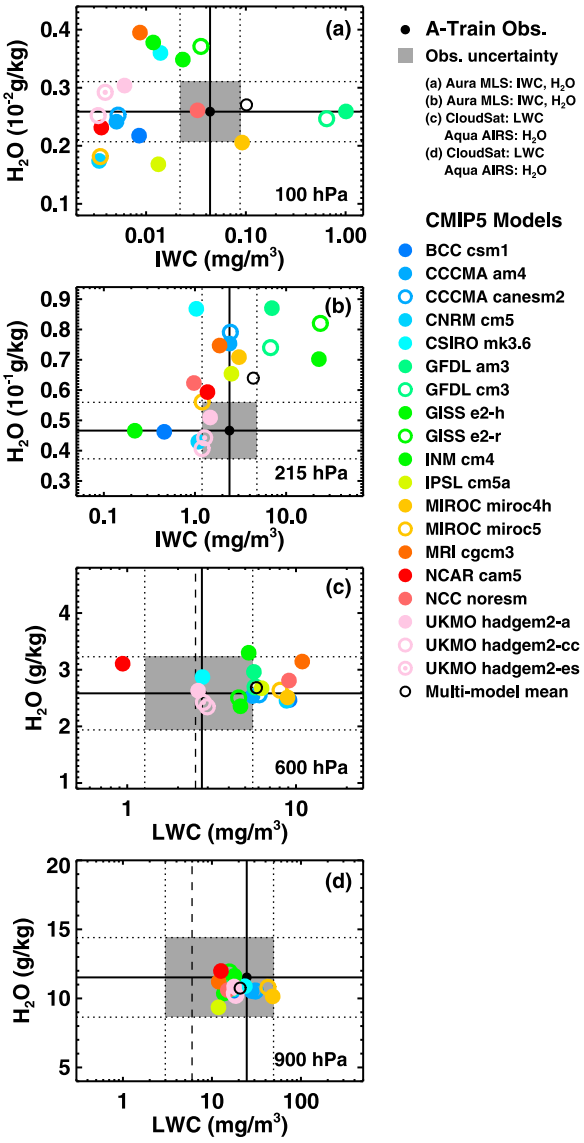

Figure 9. Scatterplots of tropical (30°N–30°S) oceanic multiyear means: H_2O versus IWC at (a) 100 and (b) 215 hPa, and H_2O versus LWC at (c) 600 and (d) 900 hPa. Black dots show the A-Train observed values and the gray area indicates the observational uncertainties. Colored dots/cycles are the values from the CMIP5 models. Black open-circles represent the multimodel means. At 600 and 900 hPa, the black dots are the CloudSat *Total* and dashed lines indicate the CloudSat *noPcp*.

[47] Similarly, we define the spatial variance scores G_v as:

$$G_v^{IWC,\,LWC} = \max\left[0, 1 - \frac{1}{n_g}\frac{\left|\ln\sigma_{mdl}^{IWC,LWC} - \ln\sigma_{obs}^{IWC,LWC}\right|}{\ln\varepsilon_{v,obs}^{IWC,LWC}}\right], \quad (3)$$

$$G_v^{H_2O} = \max\left[0,\ 1 - \frac{1}{n_g}\frac{\left|\sigma_{mdl}^{H_2O} - \sigma_{obs}^{H_2O}\right|}{\varepsilon_{v,obs}^{H_2O}\sigma_{obs}^{H_2O}}\right], \quad (4)$$

where σ_{mdl} and σ_{obs} are the standard deviations from models and observations, respectively. The uncertainty of the observed spatial variance, $\varepsilon_{v,obs}$, is the same as for $\varepsilon_{m,obs}$ discussed above and the same n_g values are also used here.

[48] For the spatial distribution performance, we simply use spatial correlations between model and observation as the scoring system:

$$G_c = \max\left[0, C_{mdl,obs}\right], \quad (5)$$

where $C_{mdl,obs}$ is the spatial correlation between the multiyear mean from a model and the multiyear mean from the A-Train.

5.2. Bivariate Metrics for H_2O and LWC/IWC

[49] As H_2O is strongly coupled with LWC/IWC, it is informative to simultaneously analyze the model performances for H_2O and LWC/IWC. This is particularly useful in the tropical tropopause layer (TTL) where the sum of IWC and H_2O is nearly constant [e.g., *Flury et al.*, 2011]. We thus use bivariate metrics (BVC) in the following sections to simultaneously evaluate the model performances for H_2O and for LWC/IWC.

5.2.1. Model Performances in Regards to Spatial Means

[50] Figure 9 shows scatterplots of H_2O versus IWC at 100 hPa and 215 hPa, and H_2O versus LWC at 600 and 900 hPa. Black dots, and horizontal and vertical lines, show the A-Train multiyear means; the gray area indicates the observational uncertainties. Colored dots/cycles are the multiyear means from the CMIP5 various models. Black open-circles represent the multimodel means. Tables 4a and 4b give numerical values for the spatial means, and for the resulting performance scores discussed below.

[51] At 100 hPa, only one model (NCC noresm) falls into the gray area. It scores 0.86 for IWC and 0.97 for H_2O. Most models underestimate the IWC amount, while the two GFDL models greatly overestimate it. The model biases for H_2O are split between positive and negative, and there is no apparent correlation between the biases in modeled IWC and H_2O. For example, GISS e2-r receives the highest score (0.9) for IWC, but zero for H_2O; GFDL am3 and cm3 perform excellently in simulating 100 hPa H_2O with scores of 1.0 and 0.84, respectively, but perform poorly in simulating 100 hPa IWC. The multimodel mean for H_2O is close to the MLS measurement, while the multimodel mean for IWC is barely within the observational uncertainty, resulting from the extremely high values from GFDL models compensating the general low biases in other models. The numerical scores clearly reflect the overall poor model performance at 100 hPa: of the 19 models, 10 have zeros for IWC and 8 have zeros for H_2O, with three having zeros in both IWC and H_2O. It should be noted that, because the MLS H_2O uncertainty at here is only 10%, any model producing 100 hPa H_2O that differs from the MLS value by ≥30% receives a zero score. Also, vertical interpolation may contribute to some of these biases: we found that different vertical interpolation schemes could change the H_2O bias by up to 20%.

[52] Model performance at 215 hPa is generally better than at 100 hPa. Five models are within the uncertainty limits of the observations: the three UKMO models, CNRM

Table 4a. Spatial Means $\overline{IWC}_{mdl}/\overline{LWC}_{mdl}$ and Spatial Mean Scores $G_m^{IWC/LWC}$ for IWC and LWC[a]

	100 hPa (MLS) 0.0438 (0.0219–0.0875) (mg/m^3)		215 hPa (MLS) 2.39 (1.20–4.78) (mg/m^3)		600 hPa (CloudSat) 2.77 (1.27–5.55) (mg/m^3)		900 hPa (CloudSat) 24.4 (3.06–48.8) (mg/m^3)	
CMIP5 Model	$\overline{IWC}_{mdl}$	G_m^{IWC}	$\overline{IWC}_{mdl}$	G_m^{IWC}	$\overline{LWC}_{mdl}$	G_m^{LWC}	$\overline{LWC}_{mdl}$	G_m^{LWC}
BCC csm1	0.00851	0.21	0.460	0.21	9.16	0.43	18.4	0.90
CCCMA am4	0.00505	0.0	2.39	1.0	5.52	0.67	27.9	0.95
CCCMA canesm2	0.00523	0.0	2.44	0.99	6.05	0.63	30.8	0.92
CNRM cm5	0.00338	0.0	1.09	0.62	8.79	0.45	18.0	0.89
CSIRO mk3.6	0.0139	0.45	1.03	0.60	2.79	1.0	23.5	0.99
GFDL am3	1.01	0.0	6.98	0.48	5.63	0.66	15.5	0.84
GFDL cm3	0.646	0.0	6.75	0.50	5.72	0.65	16.3	0.85
GISS e2-h	0.0234	0.70	22.9	0.0	4.69	0.75	17.9	0.89
GISS e2-r	0.0354	0.90	23.8	0.0	4.57	0.76	15.7	0.84
INM cm4	0.00393	0.0	0.0729	0.0	1.75	0.78	4.53	0.39
IPSL cm5a	0.0133	0.43	2.51	0.98	6.26	0.61	11.8	0.74
MIROC miroc4h	0.0918	0.64	3.04	0.88	8.91	0.44	48.2	0.75
MIROC miroc5	0.00347	0.0	1.20	0.67	8.05	0.49	42.7	0.80
MRI cgcm3	0.00868	0.22	1.86	0.88	10.9	0.34	11.9	0.74
NCAR cam5	0.00356	0.0	1.37	0.73	0.940	0.48	12.6	0.76
NCC noresm	0.0328	0.86	0.974	0.57	9.09	0.43	15.1	0.83
UKMO hadgem2-a	0.00607	0.05	1.47	0.77	2.63	0.97	17.8	0.89
UKMO hadgem2-cc	0.00330	0.0	1.20	0.67	2.98	0.97	18.5	0.90
UKMO hadgem2-es	0.00389	0.0	1.28	0.70	2.83	0.99	17.9	0.89

[a]Observed means and their uncertainty ranges are included in the column headings.

and MIROC5. Their scores are higher than 0.6 for both IWC and H_2O. INM simulates 215 hPa H_2O extremely well (scored 1.0), but significantly underestimates IWC (scored zero). The low bias of CWC in INM is persistent at all vertical levels, consistent with its low bias in IWP and LWP (Figures 1 and 2). Most models overestimate the H_2O amount at 215 hPa, and several tend to overestimate both IWC and H_2O. At this level, IWC is better simulated than H_2O: only 3 models have low skill (scored 0–0.2) for IWC, compared to 9 models with scores for H_2O lower than 0.2 (6 of them are zeros). This may suggest a poor model representation of convective detrainment and subsequent moistening of upper troposphere by detrained cloud particles [*Su et al.*, 2006a, 2006b]. The multimodel mean at 215 hPa indicates an overestimate of both IWC and H_2O.

[53] At 600 hPa, all model simulations of H_2O are generally within the observational uncertainty, while simulated LWC tend to be larger than the observed (only two models fall short). The scores for all models are higher than 0.3 for LWC and 0.6 for H_2O, with the medians being 0.65 for LWC and 0.93 for H_2O. The multimodel mean for LWC is at the edge of the maximum observation uncertainty, while the multimodel mean for H_2O closely matches the AIRS observation.

[54] At 900 hPa, model LWCs range from 4.53 mg/m^3 (INM cm4) to 48.2 mg/m^3 (MIROC miroc4h) and are all within the CloudSat observational uncertainty. Scores for LWC are better than 0.7, except for INM (0.39), due to its LWC being even smaller than the CloudSat *noPcp* value. All models perform well for 900 hPa H_2O, with scores greater

Table 4b. Model Spatial Means $\overline{H_2O}_{mdl}$ and Spatial Mean Scores $G_m^{H_2O}$ for H_2O[a]

	100 hPa (MLS) 0.259 ($\pm$0.0259)10^{-2} (g/kg)		215 hPa (MLS) 0.466 ($\pm$0.0932)10^{-1} (g/kg)		600 hPa (AIRS) 2.58 ($\pm$0.646) (g/kg)		900 hPa (AIRS) 11.5 ($\pm$2.88) (g/kg)	
CMIP5 Model	$\overline{H_2O}_{mdl}$	$G_m^{H_2O}$	$\overline{H_2O}_{mdl}$	$G_m^{H_2O}$	$\overline{H_2O}_{mdl}$	$G_m^{H_2O}$	$\overline{H_2O}_{mdl}$	$G_m^{H_2O}$
BCC csm1	0.217	0.47	0.462	0.99	2.46	0.94	10.3	0.85
CCCMA am4	0.241	0.78	0.754	0.0	2.53	0.98	10.5	0.88
CCCMA canesm2	0.253	0.92	0.791	0.0	2.56	0.99	10.5	0.89
CNRM cm5	0.174	0.0	0.430	0.87	2.45	0.93	10.7	0.90
CSIRO mk3.6	0.360	0.0	0.868	0.0	2.87	0.85	10.9	0.93
GFDL am3	0.259	1.0	0.871	0.0	2.96	0.81	11.1	0.95
GFDL cm3	0.247	0.84	0.740	0.021	2.70	0.94	10.7	0.90
GISS e2-h	0.348	0.0	0.702	0.16	2.35	0.88	11.6	0.99
GISS e2-r	0.371	0.0	0.820	0.0	2.50	0.96	11.9	0.96
INM cm4	0.378	0.0	0.466	1.0	3.30	0.63	10.4	0.87
IPSL cm5a	0.168	0.0	0.654	0.33	2.67	0.95	9.35	0.75
MIROC miroc4h	0.206	0.31	0.709	0.13	2.51	0.96	10.1	0.84
MIROC miroc5	0.00181	0.0	0.0561	0.66	2.64	0.97	10.8	0.92
MRI cgcm3	0.395	0.0	0.747	0.0	3.14	0.71	11.2	0.96
NCAR cam5	0.231	0.65	0.593	0.55	3.11	0.73	12.0	0.95
NCC noresm	0.261	0.97	0.623	0.44	2.81	0.88	10.6	0.89
UKMO hadgem2-a	0.304	0.42	0.510	0.85	2.63	0.98	10.9	0.92
UKMO hadgem2-cc	0.252	0.91	0.407	0.79	2.35	0.88	10.2	0.85
UKMO hadgem2-es	0.292	0.57	0.442	0.92	2.42	0.92	10.4	0.87

[a]Observed means and their uncertainty ranges are included in the column headings.

Table 5a. Model Spatial Standard Deviations $\sigma_{mdl}^{IWC/LWC}$ (Normalized to the Observed Spatial Standard Deviation), and Spatial Variance Scores $G_v^{IWC/LWC}$, for IWC and LWC

	100 hPa		215 hPa		600 hPa		900 hPa	
CMIP5 Model	σ_{mdl}^{IWC}	G_v^{IWC}	σ_{mdl}^{IWC}	G_v^{IWC}	σ_{mdl}^{LWC}	G_v^{LWC}	σ_{mdl}^{LWC}	G_v^{LWC}
BCC csm1	0.137	0.043	0.0949	0.0	2.88	0.49	0.384	0.66
CCCMA am4	0.117	0.0	0.869	0.93	3.42	0.41	0.701	0.87
CCCMA canesm2	0.137	0.042	0.911	0.96	3.73	0.37	0.827	0.93
CNRM cm5	0.0989	0.0	0.418	0.58	3.18	0.44	0.388	0.66
CSIRO mk3.6	0.186	0.19	0.410	0.57	1.08	0.96	0.842	0.94
GFDL am3	27.7	0.0	2.893	0.49	2.13	0.64	0.382	0.65
GFDL cm3	17.1	0.0	2.570	0.55	1.93	0.68	0.320	0.59
GISS e2-h	1.77	0.72	10.1	0.0	3.94	0.34	0.422	0.69
GISS e2-r	2.86	0.50	10.7	0.0	3.56	0.39	0.488	0.74
INM cm4	0.0666	0.0	0.0216	0.0	0.578	0.74	0.0767	0.074
IPSL cm5a	0.333	0.47	0.807	0.90	2.88	0.49	0.478	0.73
MIROC miroc4h	1.83	0.71	1.12	0.95	3.86	0.35	0.920	0.97
MIROC miroc5	0.0592	0.0	0.433	0.60	3.56	0.39	0.666	0.85
MRI cgcm3	0.222	0.28	0.674	0.81	3.84	0.35	0.221	0.46
NCAR cam5	0.0929	0.0	0.492	0.66	0.451	0.62	0.668	0.86
NCC noresm	0.744	0.86	0.244	0.32	3.09	0.46	0.567	0.80
UKMO hadgem2-a	0.173	0.16	0.499	0.67	0.930	0.97	0.562	0.79
UKMO hadgem2-cc	0.0936	0.0	0.407	0.57	0.996	1.0	0.449	0.71
UKMO hadgem2-es	0.116	0.0	0.437	0.60	0.983	0.99	0.462	0.72

than 0.7. The multimodel mean at this level is very close to the observed.

[55] Overall, the model mean IWC/LWC and H_2O averaged over tropical oceans have larger spread in the upper troposphere than in the middle and lower troposphere. Models tend to underestimate IWC at 100 hPa and overestimate H_2O at 215 hPa. It is not obvious that these two biases are coupled, or how much the models' cloud microphysics or convective schemes contribute to these problems.

5.2.2. Model Performances in Regards to Spatial Variations

[56] We now examine the degree to which the spatial variations in the multiyear means from the CMIP5 models reproduce the spatial variations in the A-Train observations over 30°S–30°N oceanic regions. Tables 5a and 5b give numerical values for the spatial variance (standard deviation) and the resulting spatial variance scores. Tables 6a and 6b give numerical values for the spatial correlation and the resulting spatial correlation scores. Subsections below discuss the model performances at each of the 4 vertical levels. Figure 10 gives Taylor diagrams for H_2O at 100, 215, 600 and 900 hPa, for IWC at 100 and 215 hPa, and for LWC at 600 and 900 hPa. Results are shown for all the 19 CMIP5 models that produce vertical profiles of H_2O, IWC and LWC.

[57] The general differences of modeled standard deviation from the observations are consistent with the differences in spatial means. From 900 hPa to 100 hPa, there is a more than 10-times increase of the ratio of modeled standard deviation for CWC relative to the observed, suggesting a

Table 5b. Model Spatial Standard Deviations $\sigma_{mdl}^{H_2O}$ (Normalized to the Observed Spatial Standard Deviation), and Spatial Variance Scores $G_v^{H_2O}$, for H_2O

	100 hPa		215 hPa		600 hPa		900 hPa	
CMIP5 Model	$\sigma_{mdl}^{H_2O}$	$G_v^{H_2O}$	$\sigma_{mdl}^{H_2O}$	$G_v^{H_2O}$	$\sigma_{mdl}^{H_2O}$	$G_v^{H_2O}$	$\sigma_{mdl}^{H_2O}$	$G_v^{H_2O}$
BCC csm1	1.53	0.0	0.476	0.13	0.671	0.56	0.846	0.80
CCCMA am4	2.55	0.0	1.36	0.40	0.872	0.83	1.01	0.98
CCCMA canesm2	2.68	0.0	1.46	0.24	0.911	0.88	1.06	0.92
CNRM cm5	0.887	0.62	0.544	0.24	0.881	0.84	0.819	0.76
CSIRO mk3.6	3.17	0.0	1.43	0.29	1.12	0.85	1.03	0.96
GFDL am3	2.18	0.0	1.54	0.10	1.26	0.65	1.00	1.0
GFDL cm3	2.20	0.0	1.16	0.73	1.02	0.98	0.877	0.84
GISS e2-h	1.34	0.0	0.993	0.99	0.754	0.67	0.951	0.94
GISS e2-r	1.63	0.0	1.28	0.53	0.881	0.84	1.10	0.87
INM cm4	5.43	0.0	0.754	0.59	1.21	0.71	0.891	0.86
IPSL cm5a	0.687	0.0	0.902	0.84	1.07	0.91	0.934	0.91
MIROC miroc4h	3.42	0.0	1.14	0.77	1.10	0.87	1.03	0.97
MIROC miroc5	2.45	0.0	0.706	0.51	1.04	0.95	0.873	0.83
MRI cgcm3	1.58	0.0	1.05	0.92	1.02	0.98	0.904	0.87
NCAR cam5	1.38	0.0	0.788	0.65	1.18	0.75	0.868	0.83
NCC noresm	1.06	0.81	0.830	0.72	1.09	0.88	0.880	0.84
UKMO hadgem2-a	1.27	0.11	0.816	0.69	1.01	0.99	0.871	0.83
UKMO hadgem2-cc	1.08	0.73	0.644	0.41	0.897	0.86	0.802	0.74
UKMO hadgem2-es	1.27	0.11	0.716	0.53	0.939	0.92	0.828	0.77

Table 6a. Model-Observation Spatial Correlation Coefficients $C^{IWC/LWC}_{mdl,obs}$, and Model Spatial Correlation Scores $G^{IWC/LWC}_{c}$, for IWC/LWC

	100 hPa		215 hPa		600 hPa		900 hPa	
CMIP5 Model	$C^{IWC}_{mdl,obs}$	G^{IWC}_{c}	$C^{IWC}_{mdl,obs}$	G^{IWC}_{c}	$C^{LWC}_{mdl,obs}$	G^{LWC}_{c}	$C^{LWC}_{mdl,obs}$	G^{LWC}_{c}
BCC csm1	0.706	0.71	0.812	0.81	0.613	0.61	0.229	0.23
CCCMA am4	0.831	0.83	0.813	0.81	0.367	0.37	0.377	0.38
CCCMA canesm2	0.728	0.73	0.784	0.78	0.371	0.37	0.336	0.34
CNRM cm5	0.613	0.61	0.830	0.83	0.661	0.66	0.143	0.14
CSIRO mk3.6	0.664	0.66	0.818	0.82	0.601	0.60	0.751	0.75
GFDL am3	0.818	0.82	0.894	0.89	0.812	0.81	0.729	0.73
GFDL cm3	0.746	0.75	0.794	0.79	0.662	0.66	0.639	0.64
GISS e2-h	0.258	0.26	0.642	0.64	−0.0294	0.00	0.479	0.48
GISS e2-r	0.241	0.24	0.677	0.68	−0.0364	0.00	0.523	0.52
INM cm4	0.581	0.58	0.492	0.49	0.507	0.51	0.227	0.23
IPSL cm5a	0.629	0.63	0.779	0.78	0.687	0.69	0.497	0.50
MIROC miroc4h	0.849	0.85	0.834	0.83	0.658	0.66	0.471	0.47
MIROC miroc5	0.694	0.69	0.865	0.87	0.759	0.76	0.384	0.38
MRI cgcm3	0.632	0.63	0.788	0.79	0.697	0.70	0.205	0.21
NCAR cam5	0.842	0.84	0.857	0.86	0.576	0.58	0.488	0.49
NCC noresm	0.592	0.59	0.814	0.81	0.645	0.64	0.434	0.43
UKMO hadgem2-a	0.677	0.68	0.831	0.83	0.620	0.62	0.636	0.64
UKMO hadgem2-cc	0.732	0.73	0.893	0.89	0.736	0.74	0.477	0.48
UKMO hadgem2-es	0.717	0.72	0.896	0.90	0.716	0.72	0.550	0.55

large model spread in the upper tropospheric clouds. Most models produce too little IWC at 100 hPa, thus their standard deviations are also smaller than the observed. The overestimates of IWC variabilities at 100 and 215 hPa by the GFDL and GISS models are manifest in their standard deviations, both having RMS much larger than one standard deviation of the observations. With other models producing low biases in IWC, the multimodel means at these two levels are near RMS of 1.0 (relative to the observed). In terms of spatial correlation, the 215 hPa IWC corresponds to the smallest inter-model differences: most models yield spatial correlations with the observation around 0.8, with the highest being 0.9 and the lowest being 0.5. This suggests convective schemes in models approximately capture the occurrence of deep convection, although the magnitudes are not well represented. Other factors, such as how much ice mass is detrained from a convective tower, how fast cirrus dissipate, could cause discrepancies among models. The modeled standard deviations and spatial correlations at 600 hPa are quite scattered in the Taylor Diagram (Figure 10). In particular, the ranges of standard deviation biases and RMS are up to 4 times of the observed. The spatial correlations are all below 0.8, with the two GISS models having negative correlations. At 900 hPa, the spatial correlations are worse than upper levels, with many models hovering around 0.4–0.5 and high (low) values around 0.7 (0.1). This clearly indicates the problems in simulating the locations of marine stratiform clouds.

[58] For H_2O, despite overall better performance than clouds at the same levels, the 100 hPa H_2O shows a dramatic departure from the observation in terms of spatial correlation: four models (two GISS and two CCCMA models) produce negative correlations with the MLS observation, reflected in the half-circled Taylor Diagram instead of the conventional

Table 6b. Model-Observation Spatial Correlation Coefficients $C^{H_2O}_{mdl,obs}$, and Model Spatial Correlation Scores $G^{H_2O}_{c}$, for H_2O

	100 hPa		215 hPa		600 hPa		900 hPa	
CMIP5 Model	$C^{H_2O}_{mdl,obs}$	$G^{H_2O}_{c}$	$C^{H_2O}_{mdl,obs}$	$G^{H_2O}_{c}$	$C^{H_2O}_{mdl,obs}$	$G^{H_2O}_{c}$	$C^{H_2O}_{mdl,obs}$	$G^{H_2O}_{c}$
BCC csm1	0.805	0.80	0.845	0.85	0.882	0.88	0.929	0.93
CCCMA am4	−0.075	0.00	0.898	0.90	0.921	0.92	0.946	0.95
CCCMA canesm2	−0.159	0.00	0.881	0.88	0.916	0.92	0.950	0.95
CNRM cm5	0.807	0.81	0.889	0.89	0.931	0.93	0.945	0.95
CSIRO mk3.6	0.569	0.57	0.890	0.89	0.888	0.89	0.961	0.96
GFDL am3	0.842	0.84	0.941	0.94	0.975	0.98	0.964	0.96
GFDL cm3	0.797	0.80	0.864	0.86	0.889	0.89	0.921	0.92
GISS e2-h	−0.152	0.00	0.738	0.74	0.800	0.80	0.893	0.89
GISS e2-r	−0.221	0.00	0.764	0.76	0.853	0.85	0.931	0.93
INM cm4	0.556	0.56	0.839	0.84	0.911	0.91	0.920	0.92
IPSL cm5a	0.494	0.49	0.893	0.89	0.894	0.89	0.911	0.91
MIROC miroc4h	0.558	0.56	0.857	0.86	0.912	0.91	0.957	0.96
MIROC miroc5	0.724	0.72	0.915	0.91	0.952	0.95	0.968	0.97
MRI cgcm3	0.807	0.81	0.809	0.81	0.833	0.83	0.889	0.89
NCAR cam5	0.789	0.79	0.913	0.91	0.975	0.97	0.955	0.96
NCC noresm	0.0383	0.04	0.867	0.87	0.878	0.88	0.924	0.92
UKMO hadgem2-a	0.892	0.89	0.857	0.86	0.935	0.94	0.963	0.96
UKMO hadgem2-cc	0.868	0.87	0.906	0.91	0.936	0.94	0.935	0.94
UKMO hadgem2-es	0.899	0.90	0.915	0.92	0.949	0.95	0.941	0.94

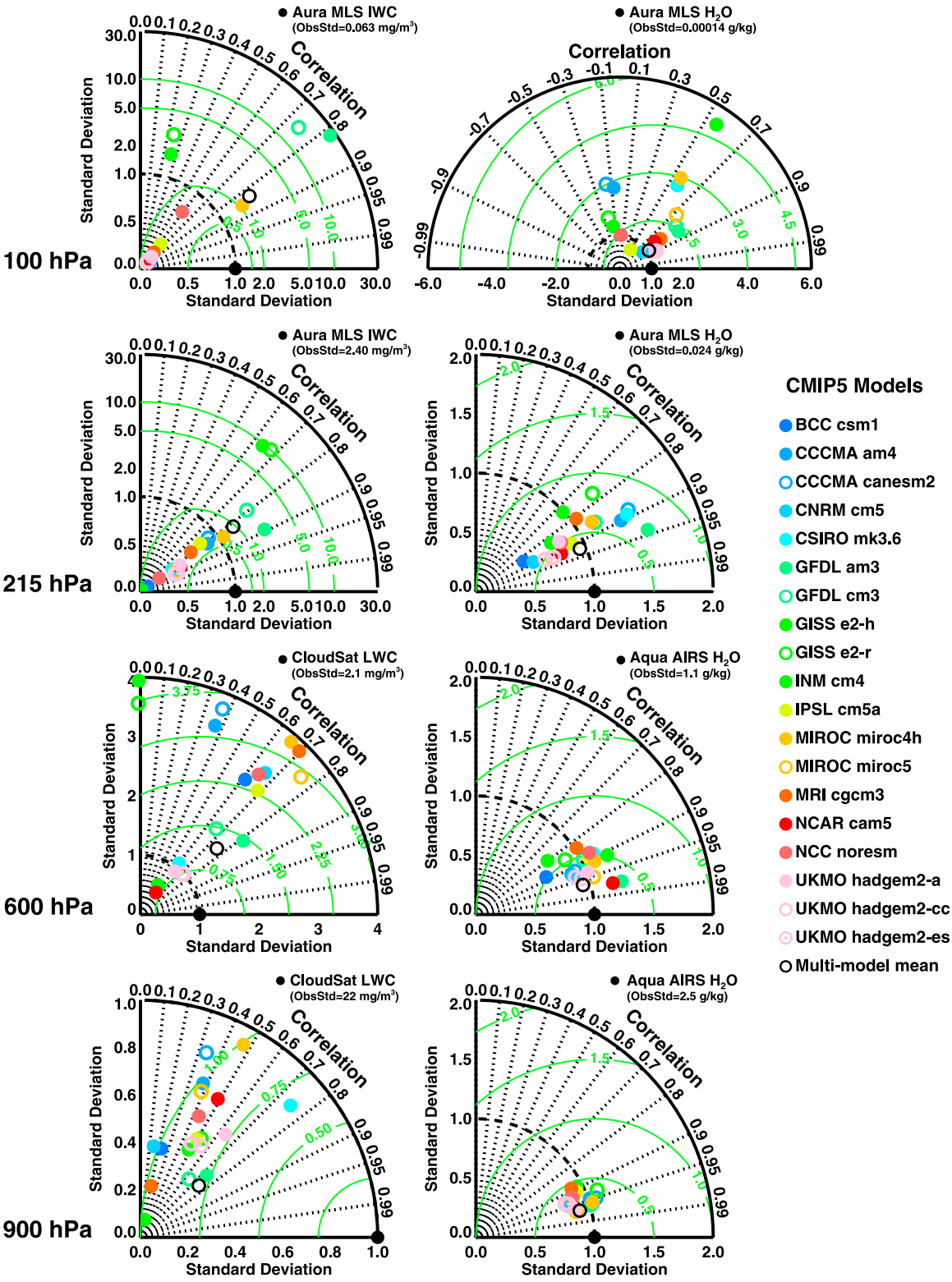

Figure 10. Taylor diagrams showing the tropical (30°N–30°S) oceanic multiyear mean performance of the CMIP5 models as compared to the A-Train observations. See text for more explanation.

quarter-circles. The RMSs from the models range from 5× to 0.5× of the observation. At the three lower levels, the modeled RMS is generally below 0.8× of the observation, and the spatial correlation is higher than 0.7. At 600 hPa and 900 hPa, the spatial correlations are more than 0.9 and the inter-model spreads are noticeably smaller than those in the upper troposphere.

[59] As the "multi-model mean" inherently smooths out individual models' spatial variations, it is not surprising the spatial variances of the "multi-model mean" are generally

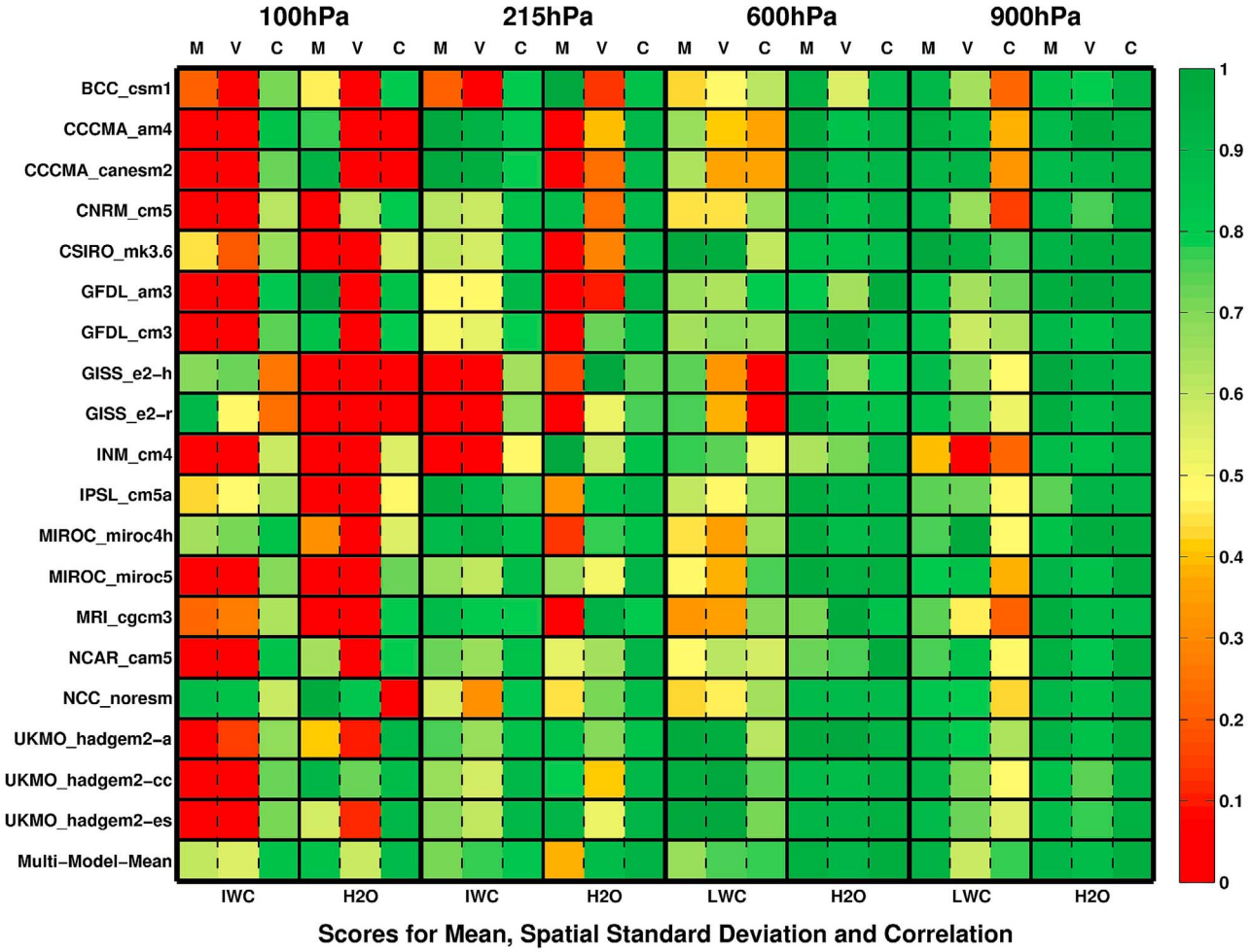

Figure 11. Color-coded summary of performance scores at 100, 215, 600, and 900 hPa. M: spatial mean performance scores G_m; V: spatial variance performance scores G_v; C: spatial correlation performance scores G_c.

closer to the observations than individual models. The spatial correlations of "multi-model mean" are also the highest among all models.

5.2.3. Overall Summary of Model Performance Scores

[60] Figure 11 gives an overall summary of all 19 models' performances in a color-coded display of each model's spatial mean, spatial variance and spatial correlation scores for all three parameters (H_2O, IWC and LWC) and all four pressure levels examined here. Although the score values are not directly comparable between IWC/LWC (clouds) and H_2O (water vapor), we find that – at all 4 pressure levels - most models simulate water vapor better than clouds.

[61] For spatial means, most models have better scores in both LWC and H_2O at 900 hPa (boundary layer) and 600 hPa (middle troposphere) than at 215 (upper troposphere) and 100 hPa (tropical tropopause layer). The simulated H_2O and IWC at 100 and 215 hPa vary greatly from model to model, indicating the large differences (and thus model uncertainty) in the parameterizations and microphysics for processes affecting high-altitude clouds. Inadequate vertical resolutions near the tropopause in the models or observations could also contribute to differences between the simulated and observed H_2O and IWC near the tropopause.

[62] For spatial variability, it is clear that models generally simulate 600 and 900 hPa H_2O (water vapor) better than LWC (clouds). Most models do not well simulate the observed variability of IWC (clouds) at 215 and 100 hPa. An interesting result is the better scores for correlation than for variance at 215 and 100 hPa, indicating that models generally simulate upper tropospheric cloud and water vapor spatial patterns (which are connected to regions of deep convection) better than they simulate the amount of spatial variation. Spatial patterns of low and mid clouds are not universally well simulated.

[63] The "multi-model mean" exhibits relatively superior performance in all aspects of metrics in Figure 11, except its score for the 215 hPa mean H_2O is below 0.5. The low score for 215 hPa H_2O reflects the fact that most models have high bias of 215 hPa spatial mean H_2O compared to the observation. On the other hand, both high and low biases exist for other quantities in the models, thus the "multi-model mean" effectively averages out the biases and achieve a better performance than many individual models. This may be comforting as the use of multimodel ensembles in climate projections is a common practice and the "multi-model mean" is generally perceived as closer to the

Table 7. Overall Scores and Ranks for the CMIP5 Models at Individual Pressure Levels

	100 hPa		215 hPa		600 hPa		900 hPa	
CMIP5 Model	Score	Rank	Score	Rank	Score	Rank	Score	Rank
BCC csm1	0.37	7	0.50	12	0.65	10	0.73	10
CCCMA am4	0.27	12	0.67	7	0.70	8	0.84	3
CCCMA canesm2	0.28	11	0.64	8	0.69	9	0.82	5
CNRM cm5	0.34	8	0.67	7	0.71	7	0.72	11
CSIRO mk3.6	0.31	10	0.53	11	0.86	3	0.92	1
GFDL am3	0.44	4	0.49	13	0.76	5	0.86	2
GFDL cm3	0.40	5	0.58	10	0.80	4	0.79	7
GISS e2-h	0.28	11	0.42	14	0.57	12	0.81	6
GISS e2-r	0.27	12	0.33	15	0.63	11	0.81	6
INM cm4	0.19	14	0.49	13	0.71	7	0.56	13
IPSL cm5a	0.34	8	0.79	1	0.76	5	0.76	9
MIROC miroc4h	0.51	3	0.74	4	0.70	8	0.83	4
MIROC miroc5	0.24	13	0.70	6	0.75	6	0.79	7
MRI cgcm3	0.32	9	0.70	6	0.65	10	0.69	12
NCAR cam5	0.38	6	0.73	5	0.69	9	0.81	6
NCC noresm	0.69	1	0.62	9	0.70	8	0.79	7
UKMO hadgem2-a	0.38	6	0.78	2	0.91	1	0.84	3
UKMO hadgem2-cc	0.54	2	0.70	6	0.90	2	0.77	8
UKMO hadgem2-es	0.38	6	0.76	3	0.91	1	0.79	7
Multi-model mean	0.72	-	0.74	-	0.84	-	0.84	-

"truth" than any single model alone, as found in previous model evaluation studies [e.g., *Gleckler et al.*, 2008].

[64] To obtain a 'first order' overall performance score for each model at each pressure level, we simply average its scores for all three variables (H_2O, IWC, LWC), and all three categories (spatial mean, spatial variance, spatial correlation) at each pressure level. Table 7 gives these scores for each model, and performance rankings in terms of it. Besides the "multi-model mean," NCC noresm has the highest 100 hPa score (0.69), followed by UKMO hadgem2-cc, and then MIROC miroc4h. IPSL cm5a has the highest 215 hPa score (0.79), followed by UKMO hadgem2-a and then UKMO hadgem2-es. Two of the UKMO models hadgem2-a/hadgen2-es also have the highest 600 hPa score (0.91), and another UKMO model hadgem2-cc ranks the second at 600 hPa, followed by CSIRO mk3.6 at third. The CSIRO model also has the highest 900 hPa score (0.92), followed by GFDL am3 and UKMO hadgem2-a.

6. Conclusions

[65] Using A-Train observations, we have assessed the simulated multiyear mean of cloud and water vapor by CMIP3 and CMIP5 models submitted for IPCC reports. For 12 CMIP5 models that have counterparts in CMIP3, we find measurable improvements from CMIP3 to CMIP5. For clouds, the GISS models (e2-h and e2-r) have improved significantly in IWP simulations. Apparent improvements in model simulations of IWP are also identified in 4 other CMIP5 models (CCCMA canesm2, GFDL cm3, MIROC miroc5, and UKMO hadgem2-a). For LWP, improvements are found in 7 CMIP5 models (CNRM cm5, CSIRO mk3.6, GFDL cm3, INM cm4, IPSL cm4, MIROC miroc5, and NCAR cam5), compared with their corresponding CMIP3 versions. For water vapor, changes in WVP from CMIP3 to CMIP5 are insignificant relative to the uncertainties in the observations.

[66] We have also examined vertical structure of CWC and H_2O produced by the 19 CMIP5 models. Both the largest spread among models and the largest differences between models and A-Train observations are at the upper tropospheric levels.

[67] We have developed a grading scheme to quantitatively evaluate model performance in simulating clouds and water vapor at different vertical levels (from boundary layer to tropopause) over the tropical (30°N–30°S) oceans in terms of spatial mean, correlation and standard deviation. Overall, we find water vapor is better simulated than clouds. Boundary layer water vapor is the best simulated, apparently because of the strong constraint imposed by SST. Tropopause layer water vapor is poorly simulated with respect to observations. This likely results from temperature biases. An analysis of relative humidity (RH) would be useful; however, RH near the tropopause is not well observed by satellites (e.g., MLS's RH has large uncertainty due to uncertainties in the temperature measurement [*Schwartz et al.*, 2008]). For spatial mean, upper troposphere ice clouds are worse simulated than lower or middle troposphere liquid clouds. In terms of spatial correlation, clouds at 215 hPa are better simulated than boundary layer clouds. Spatial variances of clouds at all levels are poorly simulated, compared with A-Train observations.

[68] Although our scoring scheme is simple, it provides a quantitative measure of the relative skills of current models in simulating clouds and water vapor.

[69] **Acknowledgments.** The NASA ROSES10 AST and COUND programs fund this project. The authors acknowledge the supports by the Aura MLS team and the Climate Science Center at the Jet Propulsion Laboratory, California Institute of Technology, sponsored by NASA. Jonathan Jiang and Hui Su are also grateful to Debbie Vane and CloudSat project for support; Tristan L'Ecuyer thanks the NASA CloudSat Science grant NAS5–99237; Mark Ringer acknowledges the support by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). We thank helpful discussion and comments from Peter Gleckler, Karl Taylor, Stephen Klein and Curt Covey of PCMDI, Lawrence Livermore National Laboratory; Veronika Eyring of Institute of Atmospheric Physics, Germany; William Rossow of City College of New York; Mark Schoeberl of Science and Technology Corporation; Brian Kahn of the AIRS team; Stephen Platnick of the MODIS team; and Melody Avery of the CALIPSO team. The three internal reviewers from CCCMA and CSIRO,

as well as the three anonymous reviewers of this journal, provided constructive comments. We are also very thankful to our colleagues from climate modeling centers across the globe, including BCC, BCCR, CCCMA, CNRM, QCCCE, CSIRO, GFDL, GISS, INM, IPSL, MIROC, MRI, NCAR, NCC, and UKMO Hadley Centre.

References

Arora, V. K., J. F. Scinocca, G. J. Boer, J. R. Christian, K. L. Denman, G. M. Flato, V. V. Kharin, W. G. Lee, and W. J. Merryfield (2011), Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, *Geophys. Res. Lett.*, *38*, L05805, doi:10.1029/2010GL046270.

Austin, R. T., A. J. Heymsfield, and G. Stephens (2009), Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature, *J. Geophys. Res.*, *114*, D00A23, doi:10.1029/2008JD010049.

Bodas-Salcedo, A., M. J. Webb, M. E. Brooks, M. A. Ringer, K. D. Williams, S. F. Milton, and D. R. Wilson (2008), Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities, *J. Geophys. Res.*, *113*, D00A13, doi:10.1029/2007JD009620.

Bony, S., et al. (2006), How well do we understand and evaluate climate change feedback processes?, *J. Clim.*, *19*, 3445–3482, doi:10.1175/JCLI3819.1.

Cess, R. D., et al. (1996), Cloud feedback in atmospheric general circulation models: An update, *J. Geophys. Res.*, *101*, 12,791–12,794, doi:10.1029/96JD00822.

Chen, W., C. Woods, J. Li, D. Waliser, J. Chern, W. Tao, J. H. Jiang, and A. M. Tompkins (2011), Partitioning CloudSat ice water content for comparison with upper-tropospheric ice in global atmospheric models, *J. Geophys. Res.*, *116*, D19206, doi:10.1029/2010JD015179.

Collins, W. J., et al. (2011), Development and evaluation of an Earth-system model–HadGEM2, *Geosci. Model Dev. Discuss.*, *4*, 997–1062, doi:10.5194/gmdd-4-997-2011.

Déqué, M., C. Dreveton, A. Braun, and D. Cariolle (1994), The ARPEGE-IFS atmosphere model: A contribution to the French community climate modeling, *Clim. Dyn.*, *10*, 249–266, doi:10.1007/BF00208992.

Derbyshire, S. H., A. V. Maidens, S. F. Milton, R. A. Stratton, and M. R. Willett (2011), Adaptive detrainment in a convective parametrization, *Q. J. R. Meteorol. Soc.*, *137*, 1856–1871, doi:10.1002/qj.875.

Diansky, N. A., and E. M. Volodin (2002), Simulation of present-day climate with a coupled Atmosphere-ocean general circulation model, *Izv. Acad. Sci. USSR Atmos. Oceanic Phys. (Engl. Transl.)*, *38*(6), 732–747.

Diansky, N. A., A. V. Bagno, and V. B. Zalesny (2002), Sigma model of global ocean circulation and its sensitivity to variations in wind stress, *Izv. Acad. Sci. USSR Atmos. Oceanic Phys. (Engl. Transl.)*, *38*(5), 477–494.

Donner, L. J., et al. (2011), The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component of the GFDL coupled model CM3, *J. Clim.*, *24*, 3484–3519, doi:10.1175/2011JCLI3955.1.

Douglass, A. R., M. J. Prather, T. M. Hall, S. E. Strahan, P. J. Rasch, L. C. Sparling, L. Coy, and J. M. Rodriguez (1999), Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft, *J. Geophys. Res.*, *104*, 27,545–27,564, doi:10.1029/1999JD900827.

Dufresne, J.-L., J. Quaas, O. Boucher, S. Denvil, and L. Fairhead (2005), Contrasts in the effects on climate of anthropogenic sulfate aerosols between the 20th and the 21st century, *Geophys. Res. Lett.*, *32*, L21703, doi:10.1029/2005GL023619.

Dufresne, J.-L., et al. (2012), Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, *Clim. Dyn.*, in press.

Eaton, B. (2011), User's Guide to the Community Atmosphere Model CAM-5.1, report, UCAR, Boulder, Color. [Available at http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/ug5_1/book1.html.]

Eliasson, S., S. A. Buehler, M. Milz, P. Eriksson, and V. O. John (2011), Assessing observed and modelled spatial distributions of ice water path using satellite data, *Atmos. Chem. Phys.*, *11*, 375–391, doi:10.5194/acp-11-375-2011.

Eriksson, P., et al. (2008), Comparison between early Odin-SMR, Aura MLS and CloudSat retrievals of cloud ice mass in the upper tropical troposphere, *Atmos. Chem. Phys.*, *8*(7), 1937–1948, doi:10.5194/acp-8-1937-2008.

Fetzer, E. J., B. H. Lambrigtsen, A. Eldering, H. H. Aumann, and M. T. Chahine (2006), Biases in total precipitable water vapor climatologies from Atmospheric Infrared Sounder and Advanced Microwave Scanning Radiometer, *J. Geophys. Res.*, *111*, D09S16, doi:10.1029/2005JD006598.

Flury, T., D. L. Wu, and W. G. Read (2011), Correlation among cirrus ice content, water vapor and temperature in the TTL as observed by CALIPSO and Aura/MLS, *Atmos. Chem. Phys.*, *12*, 683–691, doi:10.5194/acp-12-683-2012.

Gent, P. R., et al. (2011), The Community Climate System Model Version 4, *J. Clim.*, *24*, 4973–4991, doi:10.1175/2011JCLI4083.1.

Gettelman, A., et al. (2010), Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends, *J. Geophys. Res.*, *115*, D00M08, doi:10.1029/2009JD013638.

GFDL Global Atmosphere Model Development Team (2004), The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations, *J. Clim.*, *17*, 4641–4673, doi:10.1175/JCLI-3223.1.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *J. Geophys. Res.*, *113*, D06104, doi:10.1029/2007JD008972.

Gregory, D., and P. R. Rowntree (1990), A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure, *Mon. Weather Rev.*, *118*, 1483–1506, doi:10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2.

Hansen, J., A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner (1984), Climate sensitivity: Analysis of feedback mechanisms, in *Climate Processes and Climate Sensitivity*, *Geophys. Monogr. Ser.*, vol. 29, edited by J. E. Hansen and T. Takahashi, pp. 130–163, AGU, Washington, D. C.

Harrison, E., P. Minnis, B. Barkstrom, V. Ramanathan, R. Cess, and G. Gibson (1990), Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment, *J. Geophys. Res.*, *95*(D11), 18,687–18,703, doi:10.1029/JD095iD11p18687.

Hartmann, D., and D. A. Short (1980), On the use of Earth radiation budget statistics for studies of clouds and climate, *J. Atmos. Sci.*, *37*(6), 1233–1250, doi:10.1175/1520-0469(1980)037<1233:OTUOER>2.0.CO;2.

Haynes, J. M., T. S. L'Ecuyer, G. L. Stephens, S. D. Miller, C. Mitrescu, N. B. Wood, and S. Tanelli (2009), Rainfall retrieval over the ocean with spaceborne W-band radar, *J. Geophys. Res.*, *114*, D00A22, doi:10.1029/2008JD009973.

Heymsfield, A. J., et al. (2008), Testing IWC retrieval methods using radar and ancillary measurements with in situ data, *J. Appl. Meteorol. Climatol.*, *47*(1), 135–163, doi:10.1175/2007JAMC1606.1.

Holloway, C. E., and J. D. Neelin (2009), Moisture vertical structure, column water vapor, and tropical deep convection, *J. Atmos. Sci.*, *66*, 1665–1683, doi:10.1175/2008JAS2806.1.

Hoose, C., J. E. Kristjansson, T. Iversen, A. Kirkevåg, Ø. Seland, and A. Gettelman (2009), Constraining cloud droplet number concentration in GCMs suppresses the aerosol indirect effect, *Geophys. Res. Lett.*, *36*, L12807, doi:10.1029/2009GL038568.

Hourdin, F., et al. (2012), Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model, *Clim. Dyn.*, in press.

Hubanks, P. A., M. D. King, S. A. Platnick, and R. A. Pincus (2008), MODIS Atmospheric L3 gridded product algorithm theoretical basis document, *ATBD-MOD-30*, Goddard Space Flight Cent., Greenbelt, Md.

Jiang, J. H., et al. (2005), Seasonal variation of gravity wave sources from satellite observation, *Adv. Space Res.*, *35*, 1925–1932, doi:10.1016/j.asr.2005.01.099.

Jiang, J. H., et al. (2010), Five-year (2004–2009) Observations of upper tropospheric water vapor and cloud ice from MLS and comparisons with GEOS-5 analyses, *J. Geophys. Res.*, *115*, D15103, doi:10.1029/2009JD013256.

Jones, C. D., et al. (2011), The HadGEM2-ES implementation of CMIP5 centennial simulations, *Geosci. Model Dev.*, *4*, 543–570, doi:10.5194/gmd-4-543-2011.

Kim, D., A. H. Sobel, A. D. Del Genio, Y. Chen, S. Camargo, M.-S. Yao, M. Kelley, and L. Nazarenko (2012), The tropical subseasonal variability simulated in the NASA GISS general circulation model, *J. Clim.*, in press.

Kirkevåg, A., T. Iversen, Ø. Seland, J. B. Debernard, T. Storelvmo, and J. E. Kristjánsson (2008), Aerosol-cloud-climate interactions in the climate model CAM-Oslo, *Tellus, Ser. A*, *60*, 492–512, doi:10.1111/j.1600-0870.2008.00313.x.

Kubar, T. L., and D. L. Hartmann (2008), Vertical structure of tropical oceanic convective clouds and its relation to precipitation, *Geophys. Res. Lett.*, *35*, L03804, doi:10.1029/2007GL032811.

L'Ecuyer, T. S., and J. H. Jiang (2010), Touring the atmosphere aboard the A-Train, *Phys. Today*, *63*(7), 36–41, doi:10.1063/1.3463626.

Li, J.-L., et al. (2005), Comparisons of EOS MLS cloud ice measurements with ECMWF analyses and GCM simulations: Initial results, *Geophys. Res. Lett.*, *32*, L18710, doi:10.1029/2005GL023788.

Li, J.-L., J. H. Jiang, D. E. Waliser, and A. M. Tompkins (2007), Assessing consistency between EOS MLS and ECMWF analyzed and forecast estimates of cloud ice, *Geophys. Res. Lett.*, *34*, L08701, doi:10.1029/2006GL029022.

Li, J.-L., F. D. Waliser, C. Woods, J. Teixeira, J. Bacmeister, J. Chern, B.-W. Shen, A. Tompkins, W.-K. Tao, and M. Köhler (2008), Comparisons of satellite liquid water estimates with ECMWF and GMAO analyses, 20th century IPCC CMIP3 climate simulations, and GCM simulations, *Geophys. Res. Lett.*, *35*, L19710, doi:10.1029/2008GL035427.

Li, J.-L. F., D. E. Waliser, and J. H. Jiang (2011), Correction to "Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations," *Geophys. Res. Lett.*, *38*, L24807, doi:10.1029/2011GL049956.

Livesey, N. J., et al. (2007), EOS MLS version 2.2 Level 2 data quality and description document, *Tech. Rep. JPL D-33509*, Jet Propul. Lab, Pasadena, Calif.

Marchand, R., J. Haynes, G. G. Mace, T. Ackerman, and G. Stephens (2009), A comparison of simulated cloud radar output from the multiscale modeling framework global climate model with CloudSat cloud radar observations, *J. Geophys. Res.*, *114*, D00A20, doi:10.1029/2008JD009790.

Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, S. Ineson, T. Reichler, and J. Kim (2010), Analysis and reduction of systematic errors through a seamless approach to modelling weather and climate, *J. Clim.*, *23*, 5933–5957, doi:10.1175/2010JCLI3541.1.

Martin, G. M., et al. (2011), The HadGEM2 family of Met Office Unified Model climate configurations, *Geosci. Model Dev.*, *4*, 723–757, doi:10.5194/gmd-4-723-2011.

Ming, Y., V. Ramaswamy, L. J. Donner, and V. T. J. Phillips (2006), A robust parameterization of cloud droplet activation, *J. Atmos. Sci.*, *63*, 1348–1356, doi:10.1175/JAS3686.1.

Neale, R. B., et al. (2010), Description of the NCAR Community Atmosphere Model (CAM5.0), *NCAR/TN-486+STR*, NCAR, Boulder, Colo. [Available at http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf.]

Olsen, E. T., S. Granger, E. Manning, and J. Blaisdell (2007), AIRS/AMSU/HSB Version 5 Level 3 Quick Start, report, 25 pp., Jet Propul. Lab., Pasadena, Calif. [Available at http://disc.sci.gsfc.nasa.gov/AIRS/documentation/v5_docs/AIRS_V5_Release_User_Docs/V5_L3_QuickStart.pdf.]

Pierce, D. W., T. P. Barnett, E. J. Fetzer, and P. J. Gleckler (2006), Three-dimensional tropospheric water vapor in coupled climate models compared with observations from the AIRS satellite system, *Geophys. Res. Lett.*, *33*, L21701, doi:10.1029/2006GL027060.

Randall, D. A., and S. Tjemkes (1991), Clouds, the Earth's radiation budget, and the hydrologic cycle, *Global Planet. Change*, *4*(1–3), 3–9.

Randall, D. A., et al. (2007), Climate models and their evaluations, in *Climate Change 2007: The Physical Sciences Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., chap. 8, pp. 589–662, Cambridge Univ. Press, Cambridge, U. K.

Read, W. G., et al. (2007), Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H_2O and relative humidity with respect to ice validation, *J. Geophys. Res.*, *112*, D24S35, doi:10.1029/2007JD008752.

Rotstayn, L. D. (1997), A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes, *Q. J. R. Meteorol. Soc.*, *123*, 1227–1282.

Rotstayn, L. D. (1999), Climate sensitivity of the CSIRO GCM: Effect of cloud modeling assumptions, *J. Clim.*, *12*, 334–356, doi:10.1175/1520-0442(1999)012<0334:CSOTCG>2.0.CO;2.

Rotstayn, L. D., M. A. Collier, M. R. Dix, Y. Feng, H. B. Gordon, S. P. O'Farrell, I. N. Smith, and J. Syktus (2010), Improved simulation of Australian climate and ENSO-related climate variability in a GCM with an interactive aerosol treatment, *Int. J. Climatol.*, *30*, 1067–1088, doi:10.1002/joc.1952.

Rotstayn, L. D., S. J. Jeffrey, M. A. Collier, S. M. Dravitzki, A. C. Hirst, J. I. Syktus, and K. K. Wong (2012), Aerosol-induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations, *Atmos. Chem. Phys. Discuss.*, *12*, 5107–5188, doi:10.5194/acpd-12-5107-2012.

Sakamato, T. T., et al. (2012), MIROC4h—A new high resolution atmosphere-ocean coupled general circulation model, *J. Meteorol. Soc. Jpn.*, in press.

Schwartz, M. J., et al. (2008), Validation of the Aura Microwave Limb Sounder temperature and geopotential height measurements, *J. Geophys. Res.*, *113*, D15S11, doi:10.1029/2007JD008783.

Seland, Ø., T. Iversen, A. Kirkevåg, and T. Storelvmo (2008), Aerosol-climate interactions in the CAM-Oslo atmospheric GCM and investigations of associated shortcomings, *Tellus, Ser. A*, *60*, 459–491, doi:10.1111/j.1600-0870.2008.00318.x.

Soden, B. J., and I. M. Held (2006), An assessment of climate feedbacks in coupled ocean-atmosphere models, *J. Clim.*, *19*, 3354–3360, doi:10.1175/JCLI3799.1.

Stephens, G. J. (2005), Cloud feedbacks in the climate system: A critical review, *J. Clim.*, *18*(2), 237–273.

Su, H., D. E. Waliser, J. H. Jiang, J.-L. Li, W. G. Read, J. W. Waters, and A. M. Tompkins (2006a), Relationships of upper tropospheric water vapor, clouds and SST: MLS observations, ECMWF analyses and GCM simulations, *Geophys. Res. Lett.*, *33*, L22802, doi:10.1029/2006GL027582.

Su, H., W. G. Read, J. H. Jiang, J. W. Waters, D. L. Wu, and E. J. Fetzer (2006b), Enhanced positive water vapor feedback associated with tropical deep convection: New evidence from Aura MLS, *Geophys. Res. Lett.*, *33*, L05709, doi:10.1029/2005GL025505.

Su, H., J. H. Jiang, J. Teixeira, A. Gettelman, X. Huang, G. Stephens, D. Vane, and V. S. Perun (2011), Comparison of regime-sorted tropical cloud profiles observed by CloudSat with GEOS5 analyses and two general circulation model simulations, *J. Geophys. Res.*, *116*, D09104, doi:10.1029/2010JD014971.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Voldoire, A., et al. (2012), The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, doi:10.1007/s00382-011-1259-y, in press.

Volodin, E. M., N. A. Diansky, and A. V. Gusev (2010), Simulating present-day climate with the INMCM4 coupled model of atmospheric and oceanic general circulations, *Izv. Atmos. Ocean. Phys.*, *46*, 448–466.

Waliser, D. E., et al. (2009), Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.*, *114*, D00A21, doi:10.1029/2008JD010015.

Wang, J., and W. B. Rossow (1998), Effects of cloud vertical structure on atmospheric circulation in the GISS GCM, *J. Clim.*, *11*, 3010–3029, doi:10.1175/1520-0442(1998)011<3010:EOCVSO>2.0.CO;2.

Watanabe, M., et al. (2010), Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity, *J. Clim.*, *23*, 6312–6335, doi:10.1175/2010JCLI3679.1.

Waugh, D. W., and V. Eyring (2008), Quantitative performance metrics for stratospheric-resolving chemistry-climate models, *Atmos. Chem. Phys.*, *8*, 5699–5713, doi:10.5194/acp-8-5699-2008.

Wentz, F. J. (1997), A well calibrated ocean algorithm for Special Sensor Microwave/Imager, *J. Geophys. Res.*, *102*, 8703–8718, doi:10.1029/96JC01751.

Woods, C. P., D. E. Waliser, J.-L. Li, R. Austin, G. Stephens, and D. Vane (2008), Evaluating CloudSat ice water retrievals using a cloud-resolving model: Sensitivities to frozen particle properties, *J. Geophys. Res.*, *113*, D00A11, doi:10.1029/2008JD009941.

Wu, D. L., J. H. Jiang, W. G. Read, R. T. Austin, C. P. Davis, A. Lambert, G. L. Stephens, D. G. Vane, and J. W. Waters (2008), Validation of the Aura MLS cloud ice water content measurements, *J. Geophys. Res.*, *113*, D15S10, doi:10.1029/2007JD008931.

Wu, D. L., et al. (2009), Comparisons of global cloud ice from MLS, CloudSat, and correlative data sets, *J. Geophys. Res.*, *113*, D00A24, doi:10.1029/2008JD009946.

Wu, T., R. Yu, F. Zhang, Z. Wang, M. Dong, L. Wang, X. Jin, D. Chen, and L. Li (2010), The Beijing Climate Center for Atmospheric General Circulation Model (BCC-AGCM2.0.1): Description and its performance for the present-day climate, *Clim. Dyn.*, *34*, 123–147, doi:10.1007/s00382-008-0487-2.

Wu, T., et al. (2012), The 20th century global carbon cycle from the Beijing Climate Center Climate System Model (BCC CSM), *J. Clim.*, in press.

Yukimoto, S., et al. (2011), Meteorological Research Institute Earth System Model Version 1(MRI-ESM1)—Model description, *Tech. Rep. 64*, 83 pp., Meteorol. Res. Inst., Tsukuba, Japan. (Available from http://www.mri-jma.go.jp/Publish/Technical/index_en.html.)

Yukimoto, S., et al. (2012), A new global climate model of the Meteorological Research Institute: MRI-CGCM3—Model description and basic performance, *J. Meteorol. Soc. Jpn.*, *90a*, 23–64.

Zhang, Z. S., K. Nisancioglu, M. Bentsen, J. Tjiputra, I. Bethke, Q. Yan, B. Risebrobakken, C. Andersson, and E. Jansen (2012), Pre-industrial and mid-Pliocene simulations with NorESM-L, *Geosci. Model Dev. Discuss.*, *5*(1), 119–148, doi:10.5194/gmdd-5-119-2012.